DOCUMENT RESUME

ED 174 728 UD 019 667

TITLE Perspectives on the Instructional Dimensions Study: A Supplemental Report from the National Institute of Education.
INSTITUTION National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Nov 78
NOTE 80p.; For a related document, see UD 019 666

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS *Classroom Environment; *Compensatory Education; Conference Reports; *Educational Improvement; Educationally Disadvantaged; *Educational Opportunities; Educational Research; Elementary Secondary Education; Equal Education; Individualized Instruction; Program Administration; *Program Evaluation; Remedial Reading Programs; Teaching Methods
IDENTIFIERS *Elementary Secondary Education Act Title I; *Instructional Dimensions Study

ABSTRACT

In this volume the April, 1978 Instructional Dimensions Study Conference is reported. The conference focused on four areas of compensatory education: (1) opportunity to learn; (2) instructional setting; (3) planning, organization, and management; and (4) individualized instruction. An overview of the proceedings outlines and briefly discusses each of these areas. Individual papers address particular issues in each of the four areas. Papers by Gaea Leinhart and Herbert Zimiles deal respectively with the concept of educational opportunity at the classroom level and with the "engagedness" of children in the classroom. Each reviews pertinent literature on the subject of educational opportunity. G. Charlotte Kennedy analyzes teacher reported dimensions of compensatory education instructional environments in terms of "mainstream" and "pullout" settings. Roger W. Shuy argues that educational researchers have made little progress in assessing the effect of setting on education. Morton Botel describes different aspects of planning, organization, and management of selected reading programs. Finally, both Marianne Amarel and Jane A. Stallings argue that individualized instruction can be effective. (EB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174728

JUL 2 6 1979

Perspectives on the Instructional Dimensions Study

A Supplemental Report from the
National Institute of Education

November 1978

The National Institute of Education
U.S. Department of Health, Education, and Welfare
Washington, D.C. 20208

UD019667

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

Preface

The Instructional Dimensions Study (IDS) Conference and this volume reflect the cooperative effort of contractors, local school system personnel, and the National Institute of Education (NIE) staff.

Special thanks are due to Kirschner Associates, Inc., for assisting NIE in identifying teacher participants and to the National Institute for Community Development for providing important planning and support services during the conference.

Throughout the study, local school systems facilitated the IDS research effort. The teachers who participated in the conference provided many valuable insights, which were used by the authors in drafting their papers.

Michael Cohen, Joe Dominic, Nellie Santiago-Walpow, and Marcia Whiteman of NIE acted as symposium leaders for the working sessions. In this capacity, they helped clarify the principal issues and also contributed substantively to the overall conceptualization of the IDS Conference. The conference was designed by Peirce Hammond. Richard Moss was Project Officer for the conference, assisted by Karen McKee. Joy Frechtling, Peirce Hammond, Richard Moss, and Margot Nyitray read and suggested revisions for the papers. Their comments and Cathy Hodgman's special contributions to the clarity of presentation helped to better integrate the ideas of the individual papers. Each paper reflects the opinions of its author; the views expressed do not represent those of NIE or of the Department of Health, Education, and Welfare.

Contents

Page

PREFACE iii

CHAPTER I. INTRODUCTION. 1

CHAPTER II. OVERVIEW OF THE INSTRUCTIONAL DIMENSIONS STUDY CONFERENCE. Nel Noddings 5

CHAPTER III. EDUCATIONAL OPPORTUNITY 15

OPPORTUNITY TO LEARN. Gaea Leinhardt 15
OPPORTUNITY TO LEARN: ANOTHER CASE OF THERESEARCH TALE WAGGING THE DOG. Herbert Zimiles 25

CHAPTER IV. SETTING FOR INSTRUCTION 33

SCHOOL SETTING AND LEARNING. G. Charlette Kennedy 33
TOWARD FUTURE STUDY OF THE EFFECT OF SCHOOL SETTING ON LEARNING. Roger W. Shuy 40

CHAPTER V. PLANNING, ORGANIZATION, AND MANAGEMENT 49

ASPECTS OF PLANNING, ORGANIZATION, AND MANAGEMENT OF SELECTED READING PROGRAMS. Morton Botel 49
REPORTS AND COMMENTS ON THE NIE TEACHER WORKSHOPS, INSTRUCTIONAL DIMENSIONS STUDY. David Hawkins 57

CHAPTER VI. INDIVIDUALIZATION OF INSTRUCTION. 63

IS IT TRUE WHAT THEY SAY ABOUT INDIVIDUALIZED INSTRUCTION? Marianne Amarel 63
WHAT DO WE KNOW ABOUT INDIVIDUALIZED INSTRUCTION? Jane A. Stallings . . . 79

Chapter I. Introduction

Title I of the Elementary and Secondary Education Act Amendments of 1974 instructed the National Institute of Education (NIE) to conduct an evaluation of compensatory education. One component of this study was an examination of the effects of compensatory education services on children. NIE's major effort in this area, the Instructional Dimensions Study (IDS), examined teaching practices in a selected sample of classrooms to determine both whether current practices can be effective and how they might be improved. Reports on the IDS were issued in 1977 (Kirschner Associates, Inc., 1977) and sent to Congress last year (NIE, 1977).

NIE's overall mission includes two major goals. The first, promoting educational equity, was served by the reports made to Congress on the Compensatory Education Study, including IDS. In an effort to forward the second, improving educational practice, by relating IDS findings to some of the teachers who participated in the study and by considering and reporting their responses to it, NIE convened the IDS Conference in April 1978. Participating teachers were selected from among about 600 who had participated in the study. Teachers who employed a variety of teaching practices were included. Also attending the conference were nine educational researchers, selected to provide different educational perspectives. The researchers were charged with recording the conference proceedings and integrating the views of the teachers with the findings of educational research.

This volume presents the product of their efforts. Included are both an overall summary of conference proceedings and individual papers discussing issues in each of the four areas on which the conference focused: opportunity to learn; instructional setting; planning, organization, and management; and individualization of instruction.

A summary of the IDS and its findings is presented below. This summary provided the context for the conference and affords necessary background to the papers presented here.

During the 1976-77 school year, NIE examined instructional practices in 400 selected classrooms in 100 schools in 14 local educational agencies (LEAs) in the United States. The IDS assessed the effects on achievement of reading and mathematics instruction provided to compensatory education students in 1st and 3d grades in these classrooms. Participating schools were in rural, suburban, and urban areas, differed in economic levels, and were, as a whole, representative of the national ethnic composition. Selected schools also characteristically had well-implemented instructional programs, although the programs themselves varied. They ranged in approach from highly individualized diagnostic and prescriptive programs to traditional whole-class instruction.

Particular classrooms were chosen to reflect variation in the degree of individualization of instruction as measured by the use of (1) behavioral objectives, (2) individual pacing, (3) individual sequencing, and (4) diagnostic and mastery testing and prescriptive techniques. In addition, it was generally required that in selected classrooms there be teachers who had previous experience in the instructional practices studied.

Interviewers gathered data on actual classroom practices and teacher background. In order to measure the extent to which classroom instruction overlapped with the content of achievement tests, researchers analyzed the curriculum materials used by participating teachers. Videotape recording was used to examine those teacher-student interactions which would be difficult to measure through interviews (for example, frequency of individual or small-group instruction; proportion of management statements to cognitive/instructional statements). The amounts and proportions of supplemental instruction were documented by office rosters. The standard for comparison of the effectiveness of the various instructional practices was the Comprehensive Test of Basic Skills (CTBS).

The results of the IDS fall into the four areas described below:

- First, the opportunity to learn provided by relatively high amounts of instructional time and emphasis on the skills on which achievement gains were measured were important factors in high achievement gain.

- Second, in-class or mainstream instruction proved more effective than out-of-class or pullout instruction for 1st graders in reading and in mathematics and for 3d graders

in reading. However, setting was not a significant factor for 3d-grade mathematics instruction.

- Third, the overall results of the study were encouraging about the general effectiveness of compensatory instruction in well-planned and well-implemented programs.

- Fourth, the results did not show individualized instruction to be uniquely effective. Generally, individualized classrooms were neither more nor less effective than classrooms which were relatively nonindividualized.

Opportunity

The IDS examined two types of opportunities given to students. The first, opportunity to learn, can roughly be translated into instructional time. It included measures of length of schoolday, amount of regular and supplemental reading and math instruction, attendance, class and group size, proportion of students working "on task," and amount of homework.

Research evidence shows that increased instructional time is associated with increased learning, but the way time is spent is also important. Some evidence illustrates that "direct instruction" is one effective way of using instructional time.

The second type of opportunity, opportunity to demonstrate learning, is based on the premise that what is learned depends on what is taught. This rather simple assumption implies that programs will appear to be more successful when the content of the test materials is closely related to curricular content of the instructional program.

The IDS found that instructional time was an important determinant of achievement gain and that when instruction emphasized the particular skills on which achievement gains were measured, student achievement gains were especially large. Each of these findings was more pronounced for 1st grade than for 3d grade.

Setting

The question of where supplemental (compensatory) education should be provided is controversial. The two major alternatives examined by IDS were (1) in-class or mainstream instruction and (2) out-of-class or pullout instruction. The controversy about setting has two aspects.

First, some States and districts mandate the use of pullout programs because they believe that the Title I regulations require it or because they want to illustrate clearly the existence of a special program for designated students by physically isolating it from the regular classroom. The NIE National Survey of Compensatory Education found that pullout instruction predominates nationally for compensatory reading instruction. In 1st grade, pullout services were delivered to 64% of the compensatory education students receiving reading and 28% of those receiving mathematics. In 3d grade, the figures were 74% and 57%, respectively. The Title I regulations do not, in fact, require pullout programs, but they do require a special program for designated students. Since models of allowable mainstream programs have not been disseminated widely, States and districts concerned about the legality of their programs and about the risks of program audits by the U.S. Office of Education have often mandated or designed pullout programs to assure legality and minimize the risk of audit exceptions.

The second aspect of this controversy is pedagogical. Which setting, if either, benefits the Title I students more? Or, putting it negatively, does either setting tend to harm students in some fashion? Advocates of each of these two approaches make claims about the effectiveness of pullout and of mainstream instruction. Some believe that pullout instruction may make instruction easier by allowing students of approximately equal abilities and skills to be instructed together. Others believe that pullout instruction may facilitate de facto segregation or resegregation; that it may encourage tracking; that it may stigmatize the children who are pulled out; and that it may make coordinating regular and compensatory instruction more difficult.

Some believe that mainstream instruction may facilitate the coordination of regular and compensatory instruction; that it may use financial resources more effectively; and that it may allow greater positive peer influences because of the impact of "good" students on Title I students. Others believe that mainstream instruction may make it harder to meet the needs of individual students; that it may allow instructional patterns which have led to failure to be repeated during compensatory instruction; and that it may embarrass Title I students because they receive this extra instruction in the presence of their peers.

To date, research evidence is inconclusive on most of these topics, although some suggest that ability grouping ("homogeneous grouping," or tracking) is detrimental to the achievement gains of average and below-average students. The question of effects of setting on achievement is a separate but related question. It is possible, for example, to use ability grouping in both mainstream and pullout settings.

The IDS found that 1st graders gained most in mainstream settings in reading (13 months or 17 percentile points vs 12 months or 11 percentile points) and in mathematics (13 months or 20 percentile points vs 11 months or 7 percentile points). Third graders gained most in mainstream reading (10 months or 13 percentile points vs 7 months or 8 percentile points), but mathematics gains were equal in the two settings (12 months or 17 percentile points).

Planning, Organization, and Management of Compensatory Instruction

The research results are generally encouraging about the effectiveness of compensatory education programs examined in this study. Overall, students in grade 1 made average gains of 12 months in reading and 11 months in math during the 7 months between fall and spring testing as measured by the CTBS. Third graders gained 7 months in reading and 12 months in math on the CTBS. These gains exceed those reported in the most positive recent evaluations of compensatory instruction. Contributing to these results may be certain special features of the selected districts and classrooms: districts were nominated because they had well-implemented programs, and classroom teachers were generally included only if they were experienced in giving a particular type of instruction (ranging from individualized diagnostic and prescriptive instruction to more traditional whole-class instruction).

Although these classrooms were not special demonstration projects, they were better planned, organized, or managed than is typical of classrooms serving similar students. While teacher background, including teaching experience, formal education, and amounts of various kinds of recent inservice training, was not related to achievement gain within the study, there still may be aspects of teacher background (such as organizational skills) which distinguish the study sample as a whole from teachers in general.

Individualized Instruction

The IDS examined individualization of instruction for two reasons. First, Congress expressed explicit interest in it. Second, there was a substantial amount of research work indicating that such features of individualized programs as:

- clearly stated behavioral objectives which take individual students in small steps through curriculum materials;
- instruction which bears directly on the objectives;
- structured, sequential instruction;
- high-intensity, closely supervised instruction;
- individual diagnosis and prescription;
- individual or small-group instruction; and
- flexible grouping

were often associated with achievement gains greater than those of programs lacking such features. Further, many educators and educational researchers currently believe strongly in the effectiveness of "direct instruction," for example, instruction characterized by these features.

Study findings did not, however, show individualized classrooms to be uniquely effective. The IDS results showed substantial gains in reading and mathematics achievement irrespective of the presence or absence of individualization as defined for this study.

Chapter II. Overview of the Instructional Dimensions Study Conference

Nel Noddings
Stanford University

In 1974, Congress directed the National Institute of Education (NIE) to examine the purposes and operation of Title I of the Elementary and Secondary Education Act of 1965. The Congressional mandate to NIE has been characterized in terms of two principal questions: How effective are compensatory education programs in meeting their basic purposes? What changes might be made to improve the effectiveness of compensatory education programs?[1]

In attempting to answer these questions, NIE developed a strategy for research that involved an examination of four major areas: funds allocation, service delivery, student development, and program administration. The Instructional Dimensions Study (IDS) was designed as part of the study of student development. The IDS Conference was convened to supplement the information presented in the study and to aid policymakers in interpreting that information.

The conference brought together 40 teachers who had participated in the IDS and invited them to discuss the study's findings and to contribute information and suggestions they had gained in their own professional work with Title I. As NIE Deputy Director Michael Timpane pointed out in his opening remarks, the conference marked another significant effort in NIE's continuing campaign "to improve the practice of education as art, science, and profession." Researchers, he said, could say quite a lot about education as science, but teachers were needed to describe and suggest improvements in education as art and profession. His remarks proved correct. The views expressed by the conference teachers differed dramatically from those of researchers. But the difference provides a complementary view, one that fills out the scientific skeleton of research, and not a contradictory one.

In this overview, our main purpose is to describe the conference itself. To do that lucidly, however, something must be said about the IDS. For a fuller description, the reader is again referred to The Effects of Services on Student Development.

In designing the IDS, researchers took into account two well-known facts about recent research: first, global assessments of compensatory education programs have not been very useful--partly because of ambiguities in the interpretation of their findings, and partly because they have failed to establish any connection between success and particular program features which might have produced that success. Second, not all desirable outcomes are equally measurable. Instruments exist, for example, by which we can measure "self-image" and attitudes, but there is considerable disagreement over what the instruments actually measure, and the instruments frequently display a lack of sensitivity and discriminatory power. Hence, IDS researchers decided to narrow the focus of the study and to concentrate upon trying to find connections between particular program features and achievement in mathematics and reading. Adequate measurement instruments exist for both areas, and it is generally agreed that both subjects are crucial in scholastic development.

A major difference in priorities between researchers and teachers became obvious immediately. Researchers, in order to do "good science," must strip away complexity and attempt to control variables and establish hypothetical connections. Teachers, looking at the same situation, embrace the complexity and choose for their area of concentration features of the program which seem to be the murkiest scientifically. Interestingly, as Botel and Hawkins point out in their papers, teachers think that state and district policymakers also oversimplify. This trend toward what teachers see as oversimplification is partly a product of the accountability movement and its functional dependence on behavioral objectives. Thus, while a district guide may state its aims for reading in terms of specific decoding skills and vocabulary attainment, teach-

[1] A description of the history and purposes of NIE's efforts to examine Title I programs may be found in The Effects of Services on Student Development, September 1977, an NIE publication available from The National Institute of Education, Washington, D.C. 20208.

ers are likely to aim for voluntary and self-selected reading, for pleasure in silent reading, and for social outcomes resulting from shared experiences in reading. Are teachers wrong to pursue such aims? Should they concentrate on the more specifically stated and narrowly focused objectives? Are researchers wrong in their effort to circumscribe and control? Probably the answer to all three questions is no. But it must be kept in mind that sharply contrasting views follow from these very different sources.

The IDS focused on selected reading and mathematics programs in 1st and 3d grades, and it concentrated on two questions: Is compensatory instruction more effective in mainstream or pull-out settings? What classroom procedures are related to significant student achievement? The study, then, consists of four major areas: (1) opportunity to learn, (2) setting for instruction, (3) planning, organization, and management, and (4) individualization of instruction.

The IDS Conference was organized into four workshops treating these topics. From this point on, this summary is also organized around the four topics, and further description of the definitions, variables, hypotheses, interpretations, etc., will be undertaken in each of the subsummaries.

Opportunity to Learn

"Opportunity is the chance of obtaining, or the access one has to, a valued resource." (See Leinhardt's paper, p.15.) The concept of opportunity, as Leinhardt states, has been an economic and political one. As an educational concept, it holds some promise, but it also introduces some difficulties. Consider, for example, a hypothetical case of students who are exposed to an acceptable number of minutes of instruction in subject X but fail to attend to the instruction effectively because of, say, language difficulties. How shall we measure their opportunity to learn? Or, suppose children spend a significant portion of their time, as they sometimes do in highly individualized programs, doing the wrong thing--practicing their errors, failing to follow printed instructions, using a faulty heuristic. How much of their time should be counted as "opportunity to learn"?

It is clear that a concept of opportunity which uses amount of instructional time (either allocated or engaged) presupposes a certain level of proficiency on the part of the teacher: that the teacher will choose methods at least reasonably likely to be effective with given students, that the teacher will be able to supervise closely enough to ensure that "practicing errors" will not be a common occurrence. The IDS did, indeed, concentrate its attention on reasonably proficient teachers and accordingly successful outcomes. Hence, one can reasonably ask: What is the effect of increasing instructional time on student achievement?

The IDS included as variables in its measurement of "opportunity": class size, minutes per week of instruction, time students spend on task, and overlap between the curriculum (things actually taught) and test content. The last variable, "overlap," is interpreted as a measure of "opportunity to demonstrate learning." For all groups of students, overlap correlated positively with achievement.

Teachers at the conference were asked how much time they scheduled for reading and math and what sorts of strategies they used to increase instructional time when they thought it was necessary to do so. Many replied that time for reading and mathematics was mandated by State or district. Therefore, allocated time for these subjects was fixed. But it turned out that for many of these teachers, reading instruction pervaded the day's instructional activities. In social studies and science, for example, teachers spent time on specific reading instruction. The slogan "Every teacher is a reading teacher" was taken seriously by the conferees. It is therefore difficult to determine exactly how much time is spent on reading instruction in any given situation.

The difference in viewpoint between teachers and researchers (and, say, a district's behaviorally stated goals) makes it difficult for the careful researcher to pinpoint "minutes of instructional time." The conference teachers, for example, took time from "actual reading" to allow students to express themselves, share experiences, and choose their own learning paths. To the teachers, these activities were essential and integral parts of reading instruction; to the researchers, they were "time out" from direct instruction in reading skills. Thus, there was a basic conflict between teachers and researchers over both quantity and quality of "instructional time" and "engaged time." (See Zimiles, p. 25 and p. 26.)

Teachers generally rejected suggestions that an increase in instructional time be achieved by lengthening the schoolday or school year, by taking time from other subjects, or by increasing the pace of instruction. (See Leinhardt, pp. 16, 17, 18.) Indeed, they were concerned about the fact that Title I children often missed social studies, science, or music in order to receive supplemental instruction in math and reading. To prevent this "missing out," some teachers even used the time when Title I children were "pulled out" for enrichment in the areas of reading or math with their regular students. In these special circumstances, the scientific conditions are fur-

ther confounded. Can the Title I children be said to receive "supplemental instruction" if other children are also receiving special additional instruction? And, if children receive very particular help in reading associated with social studies and science, is it not possible that Title I children are actually sometimes missing a vital segment of the instructional package? It should be possible to work out some sort of scheme by which Title I children with a certain mix of skills would study the regular social studies and/or science and those with another mix receive continued instruction in general decoding skills. At the present time, our knowledge appears to lack the fine tuning required for such a scheme.

The conference teachers named organizational matters as a major factor decreasing instructional time. Motivating, disciplining, listening, sharing, and "loving" were all considered integral to the instructional process. But testing, moving large numbers of children here and there, recording progress on special cards, writing out prescriptions--the vast sum of clerical and logistical tasks--were seen as tasks in conflict with the instructional process. Many teachers seemed to have at least partially resolved this problem with the help of sympathetic administrators by constantly paring away at the mountain of paperwork. Those with less administrative support felt that the superficial appearance of accomplishing something had higher priority in their schools than did actual contributions to students' achievement and well-being.

In the matter of pacing and sequencing, we find an interesting anomaly. Leinhardt cites a considerable body of research on "timing" and "pacing," but teachers seem to go at the matter intuitively or traditionally. Thus, math and reading are taught when "long" blocks of time can be counted on. One "moves on" when students show signs of being ready to do so. Indeed, discussion at most of the Opportunity Workshops barely touched on matters of sequencing and timing, focusing, rather, on strategies for increasing instructional time or on testing as part of "curriculum overlap." Teachers might profit from responsible suggestions and reliable information about timing and pacing.

An interesting feature of the talk about instructional time centered on the problem of keeping children on-task. Teachers found one-to-one strategies relatively ineffective with Title I students because children are left to work by themselves for long periods of time while the teacher works with other individuals. They found it easier to keep children on-task in small groups where questions can be asked and answered collectively and where peer support can keep things going when the teacher moves to another group. Again, teachers emphasized the socializing effects of small groups and the opportunities for children to be recognized and helped by their peers. Teachers grouped and regrouped, struggling to find a balance that would permit a high degree of personalization together with the necessary level of monitoring. Too large a group meant a sacrifice in personalization and a level of monitoring aimed at the lower half of the group; too small a group meant many children neglected and an increase in random behavior.

During the discussion on curriculum overlap, it became clear that teachers had a host of gripes about testing. (See Zimiles, pp. 25 and 29.) Sometimes there was so much testing that it invaded time teachers saw as instructional time. Sometimes too much emphasis was placed on tests and not enough on features of the curriculum which resist accurate measurement. And sometimes tests incorporated items with which children were unfamiliar because of their geographical location or social status. Arizona children, for example, do not see rivers of the Pennsylvania sort, and Florida children are generally unfamiliar with snow gear.

Most of the teachers recognized the potential value and necessity of standardized testing but felt that committees of teachers representative of various geographical locations should be involved in the process of constructing test items. Teachers, informed of the concept being tested, could suggest forms for the particular items that would not prejudice the performance of youngsters from their area.

Finally, there was an important question about the desirability of complete curriculum overlap. Surely, it is obvious that people are likely to test better on material they have specifically learned than on material they have not addressed directly. But teachers pointed out that many of the abilities we value most, such as critical thinking, general problem solving, and aesthetic and moral sense, cannot be taught directly. Yet we feel that these abilities are largely acquired and enhanced in sensitive and enduring teaching/learning relationships. We want to teach for these outcomes, but we cannot teach them directly. It appears that teachers want tests that will measure these important effects, or, if such tests are not available, they want policymakers to understand that some vitally important educational goals are not being assessed by the tests that are presently administered.

Setting for Instruction

The workshop on setting induced some of the liveliest debate of the conference. It also aroused argumentation among the recorders, and one can

see evidence of that arousal in the papers of this volume.

The question addressed by the IDS was this: Is compensatory instruction more effective in pullout or in mainstream settings? Pullout instruction was defined as supplemental instruction delivered to students outside the regular classroom. Mainstream instruction was defined as supplemental instruction delivered within the regular classroom. The IDS findings indicated that "mainstreaming" was more effective in 1st-grade reading and mathematics and 3d-grade reading, and that the settings produced about equal results in 3d-grade mathematics.

Controversy arose first over the terms themselves. "Mainstreaming," as an educational term, is used primarily in the domain of special education. To "mainstream" an educationally handicapped child has meant to place him/her in a regular (as opposed to special) class. Some children might be "mainstreamed," for example, in mathematics, others in social studies. For the rest of the day, these children would return to their special class to study with their specially trained teacher. In recent years, "mainstreaming" has meant returning large numbers of children who have been assessed as educationally handicapped (EH) to regular classrooms. This practice has caused heated controversy. Those favoring "mainstreaming" claim that it removes the stigma of EH classification; those objecting to mainstreaming claim that the practice ignores all we have learned about teaching EH students and destroys a program which is just beginning to bud into a science.[2]

To use the word "mainstream," then, was to borrow trouble. But there are further problems. Roger Shuy accurately reflects the teachers' questions when he asks what the vital difference is between pullout and mainstream settings.

The claims usually made in favor of mainstreaming are these: children are likely to feel that they are "different" if they are "pulled out" for special instruction; children often miss important subjects such as science and social studies when they are pulled out; mainstreaming is less expensive than pullout; and finally, mainstreaming is just as effective as pullout. The conference teachers offered arguments to counter each of these claims. (For a more complete summary of the arguments on each side, see Kennedy's paper.) Let's consider some of the teachers' arguments.

[2]Some of this debate can be reviewed in the summary of the NIE Curriculum Development Conference, November 1976. The summary is available from NIE.

Perhaps children feel stigmatized when they are "pulled out" of their regular classrooms for special instruction. Most of the conference teachers felt that this was not a large problem. Sometimes, they said, children did feel like "dummies" when they were pulled out, but more often they knew themselves that they were having academic difficulties and they appreciated the special help. Sometimes whole classes needed counseling on the matter, and such counseling constituted a learning experience for all the children. (See Kennedy, p. 34.) Further, as teachers pointed out, grouping takes place even in mainstream situations, so it is mainly a matter of emphasis.

To assess the arguments here is very difficult. Again, borrowed trouble creeps in. Arguments for and against "tracking" have raged for years, and pullout looks suspiciously like tracking. Many school systems have abandoned tracking, albeit with some reluctance, because it has seemed to them that its pernicious side effects (damage to the self-image of lower track students and rabid competition among upper track students) outweighed its advantages (efficiency in teaching and student intellectual compatibility). Given this background, it is small wonder that feelings are aroused on the issue. But it must be kept in mind that grouping is an almost universally accepted practice, and, provided it does not segregate children for extended periods (e.g., whole days in a given year, a whole year in a given subject), it is not considered "tracking" but a responsible and respectable arrangement for instruction. Most of the conference teachers were involved in pullout programs, and they said that stigmatization was not a problem unique to pullout. It occurs at times in mainstream programs as well. It is not setting that aggravates or relieves the feeling of being different, but the sensitivity or insensitivity of teachers working with the Title I students.

Do children miss something when they are pulled out? This is another tough question. Shuy points out that, by and large, the conference teachers seemed not to be greatly concerned about that possibility. It is recorded, however, that some teachers felt so strongly about the potential loss that they provided enrichment for their regular students and "saved" the social studies, science, or whatever, for times when the children were together. Others, some Title I specialists, spoke of providing experiences in these subjects for Title I children within their special instruction. These experiences, it must be noted, were relatively rare.

For the most part, a pragmatic attitude prevailed among the teachers. It is better, they seemed to feel, to miss a little something now

than to miss out on education entirely. One must learn to read and to do mathematics.

Is mainstreaming less expensive than pullout? Under most definitions, the answer to this seems to be yes. It requires less space, can sometimes be accomplished with aides, and is easier to coordinate. But one must look hard at actual costs, and to make a sensible decision on this criterion, we need more than simple cost figures and achievement figures; we need a cost-effectiveness analysis. Some mainstream programs are very expensive. Are these also the most effective? Some pullout programs are relatively inexpensive. Are these the least effective? Again, it is not easy to make a policy decision on the basis of the IDS findings.

Can mainstreaming do whatever pullout does and do it just as effectively? Here we have, in essence, the question the IDS set out to investigate. The results of the IDS are ambiguous, indicating that, perhaps, pullout becomes more effective as children get older. But one must be very wary of drawing that conclusion even tentatively. Perhaps the whole thing reverses again by grade 5, when children have acquired an even greater social sense and sense of self-within-group.

If we look at just these four points and assume, for the moment, that we have been able to classify instructional situations properly as pullout or mainstream, we are left with a dilemma. We have an ideological problem: Should we separate children and risk damage to self-image? Should we ignore potentially powerful methods for working with learning disabilities and keep children together? We have an economic problem: How can we achieve the best result at a justifiable cost? We have a scientific problem: What are the results? How are they related to the conditions that produce them? And we have a philosophical curriculum problem: What subjects are most important? Which can be missed with little trauma?

There is an additional problem of definition. How do we identify mainstream classrooms and pullout classrooms? Descriptions of all sorts of mixed modes emerged in the conference. (See the papers by Kennedy and Shuy.) There were mainstream classrooms in which a special teacher came in and the regular teacher left. (Someone labeled this situation "pull in.") There were pullout situations in which the regular teacher accompanied her class to a Title I lab where she taught the regular students and a Title I specialist taught the Title I students. There were some classes entirely composed of Title I students.

The main point to be made here is one made by several of our writers. The terms "mainstream" and "pullout" are poorly defined and tend to mask what are surely more important variables, variables which must be defined in areas such as the physical quality of classrooms, emotional quality of class gatherings, quality of classroom discourse, quality of peer interaction, mobility, stability, and provision for multicultural experience.

Further, setting may very well interact with other variables such as administrative or parental support, and so mainstreaming may be better for one community and pullout better for another. Where teachers favor one setting over the other and work effectively in it, there seems no compelling reason--on the basis of research--to force a change.

Planning, Organization, and Management of Compensatory Instruction

The IDS undertook an examination of highly successful compensatory situations. The teachers' comments helped to clarify why these programs were so successful. Reading the study itself, we might have thought mainstreaming had the edge on pullout as a successful instructional setting; listening to the participants, we were forced to believe this is not the case. Something else, something more critical, underlies success in both settings. Similarly, time on-task is certainly important, but how is an increase of on-task time achieved? How do teachers, successful teachers, keep students happy and healthily task oriented?

The workshops on planning and organization suggested aspects of successful programs. First, the conference teachers certainly met the IDS criterion that only stable programs be included in the study. Most of them had worked for some time in Title I programs and were well acquainted with the aims and daily operation of their own systems. There was definitely stability, and, perhaps more important, there was continuity.

But there was also continuous refinement. More than anything else at the conference, one was impressed with the resilience and flexibility of the teacher representatives. They molded their programs on yearly, monthly, and daily bases.

First, they modified state and district objectives. Although the district might state its goals for Title I children in terms of so many units of progress in reading and math, conference teachers wanted happy, caring, self-disciplined children, voluntary readers, and competent problem solvers. They spoke again and again of the importance of self-image, not just in its intuitively obvious connection to cognitive growth, but as a humane

goal important in itself. It was clear that they would have spent time enhancing self-image even if that goal were not connected to cognitive growth. Further, they could describe what they did to enhance self-image. Teachers said that some children need to be touched. Others need to be listened to; they come from large, boisterous families and they never get a chance to talk. Some kids need to be recognized; they can't get needed recognition academically so they want to try to get it some other way. Teachers made efforts to recognize talent, even in such things as "clowning," when that recognition made a child's day.

None of this should be taken to mean that the conference teachers leaned toward a Summerhillian permissiveness with respect to studies. On the contrary, they were deeply concerned about academic growth. They adopted a teaching/parenting role in which they gave both specific instruction and the kind of cultural support we think of as typical in the well-educated home. Indeed, in his summary of this portion of the conference, David Hawkins surmised that a large portion of the recorded success of conference teachers might be due to their "acquaintance with individual children," and this possibility deserves study.

Second, they adapted curricular materials with some freedom. It is very hard to say how scientific and accurate with respect to subject matter these adaptations are, because we heard very little that was specific. We heard nothing, for example, about the detailed use of Cuisenaire rods or other math manipulatives or about specific programs for increasing decoding skills or vocabulary, but we heard about changing, modifying, and refining. It seemed that instructional mode was changed most freely. Teachers retained materials themselves and were grateful for a multiplicity of them, but they sometimes resisted prepackaged directions for their use. One teacher, for example, spoke of her "adaptation" of Project Plan. When pressed for the details of her adaptation, she said, "Well, mostly I threw it out." She had used the materials, but she constructed her own sequence and methods of instruction.

This is another area requiring serious study. These are successful teachers, and they are apparently doing something right. There remains the possibility, however, that they could do some things even better with the help of specialists who know the materials. One should add that curriculum specialists might also learn something from watching the uses to which their materials are put in successful classrooms.

Third, conference teachers showed a high degree of social responsiveness and resilience. They had ways of "getting around" people and situations they found hampering. They wanted aides, but they preferred to choose their own. They wanted to work with colleagues, but, again, they were pleased when working relationships could be formed by choice and mutual satisfaction. They saw principals in supportive roles; a "good" principal supported his or her competent teachers, but the teachers did the initiating in instructional matters.

Shuy notes in his paper that teachers display a "two cooks in the kitchen" syndrome when they are forced to work together and that perhaps they need to learn how to work more closely in teams. But it is possible that we are seeing here another facet of the science/art aspects of teaching. Good teachers want to plan together, speculate over matters together, evaluate together. But they want to perform alone and over the whole range of child activity. They do not want to be assembly line workers. Some teachers even enthusiastically endorsed the notion of staying with children for several years and this, too, is a matter for further study. Surely, what gives a teacher great satisfaction in teaching is likely to be correlated with his or her success.

Fourth, and finally, teachers felt strong enough to challenge, theoretically, administrative systems of recordkeeping, but they often felt frustration in trying to bring about actual changes in this area. As both Hawkins and Botel emphasize, these teachers feel that they know their students. Keeping track of every bit of progress that children make "on little cards" is an arduous clerical task that seems unnecessary. These teachers know (intuitively? by acquaintance?) where their students are, what they need next, and what they are likely to do with assigned tasks. The records that a machine must keep, in computer-assisted instruction, for example, a sensitive teacher need not keep. The saved time can be used in pupil contact. Again, this is a matter requiring a closer look. The recordkeeping may be serving a purpose that teachers fail to see. It may, for example, induce the teachers to reflect upon the effects of their teaching and to plan more carefully. Or, perhaps, teachers who are free of it do even better. We can only say that teachers respond to recordkeeping negatively. Most have found shortcuts.

In summary, it would seem that successful teachers value their autonomy in planning, performing, and evaluating. They see themselves as initiators and place administrators and parents in supportive roles. Typically, they spend a lot of time (3 or 4 hours daily) in planning. Marking papers and recording results are, again typically, less favored tasks. Can something be done to enhance the sense of artistry in giving feedback to

students? Is there a way that teachers could gain more satisfaction from the process of correcting, reassigning, and refining student work? The task requires, perhaps, enhancing the teacher-as-scientist and suppressing the teacher-as-artist. Is there an optimal balance?

Individualization of Instruction

The reported findings of the IDS--that individualized programs were effective but not necessarily more effective than other modes--caused some initial consternation among conference teachers. Many of them just could not believe that the enormous effort they put into "individualizing" had not paid off handsomely. As the conference proceeded, these uncomfortable feelings abated, largely because it became clear that the characteristics teachers associated with the variables--and clearly valued highly--were not always the ones researchers had used to describe the variables.

The IDS used four characteristics to identify highly individualized programs:

(1) Use of behavioral objectives

(2) Individual pacing

(3) Individual sequencing--interpreted as the "existence of alternative learning paths"

(4) Use of diagnostic/prescriptive methods

It is quite clear that teachers interpreted "pacing," "sequencing," and "diagnostic/prescriptive" very differently from the researchers.

Stallings describes in some detail how teachers looked at "pacing" and "diagnosis." In both cases, teachers interpreted the terms much more broadly than researchers. Teachers, for example, saw silent, self-selected reading as a case ipso facto of self-pacing and, therefore, a practice to be rated positively in a search for individualization. Researchers did not look at "self-selected reading" but concentrated, instead, on teachers' efforts to guide or to pace students through a predetermined arrangement of subject matter, usually a set of hierarchically arranged skills and subskills. Similarly, teachers broadened the concept of diagnosis to include a search for learning styles and preferences, emotional needs, and social strengths. When these teachers said, "I diagnose the child's problems," they did not mean that they pinpointed a child's difficulties with respect to a particular computational or decoding skill. Almost surely, they engaged in instructional diagnosis as well as global diagnosis, but at what level of proficiency and with what specific techniques we really do not know.

Many teachers, while interpreting particular variables and indicators associated with individualization more broadly than researchers, perceived individualization itself as working with individual students. Both Amarel and Stallings note this perception. In cases where teachers were not sold on individualized instruction, it was often the case that they rejected the one-to-one instructional mode. As we noted in the section on planning, teachers learned that working with individuals leaves too many other students unsupervised and often unproductive. Stallings gives support to the teachers' intuitive assessment of the one-to-one arrangement by citing research studies which also conclude that one-to-one is not an instructional arrangement highly correlated with achievement.

Here, as in many other complex situations under discussion, it must be noted that assessment of instructional arrangements across settings may be a mistake. Amarel points out that there may be interactions--among instructional arrangements, number of instructional personnel, personal preferences, student styles, administrative and parental support, physical space, and so on. We must be careful, then, in stating that one-to-one is not an effective instructional mode. If every student could receive one-to-one instruction as, say, Bertrand Russell did in his childhood, we might all opt for tutoring. But under the usual Title I circumstances, in which one teacher must work with many students whose needs are great and whose self-help techniques are not well developed, working with individuals seems to be neither efficient nor effective.

On the topic of "sequencing," a familiar difficulty appeared. Sequencing was interpreted at the conference workshops to mean the existence of alternative learning paths. Teachers seemed to interpret this variable very broadly. Under a broad interpretation, almost every teacher could properly claim that she made provision for alternate paths to learning; that is, if the learning under discussion is itself broad, e.g., "learning to express oneself in a group," "learning to read silently," the paths available are almost limitless in number. But we heard very little about the sort of sequencing that interests researchers or those of us engaged in rigorous programs of curriculum design. Do teachers take account of various existing models of computation, such as methods of subtraction? Do they attempt to steer some students to one model and others to another on the basis of observed performance factors and information about the theoretical assumptions underlying the models? Are they careful to sequence forms of computation according to information we have about inherent difficulty levels, e.g., $a \cdot b = \square$ before $a \cdot \square = c$, and $a \cdot \square = c$ before $\square \cdot b = c$? Do they

depend on preconstructed curriculum materials for this sort of sequencing? Are they aware of what is known in this area? The answers to these questions are by no means clear. At the global level, teachers are keenly aware of individual differences and preferences and seem to be constantly modifying, trying out, adding, and deleting both materials and methods. At the specific level, we need more information about teacher proficiency.

Throughout the discussions on individualization, it was clear that these teachers used their judgment freely yet responsibly. In deciding which children were entitled to Title I instruction, they used the required tests but felt free to add or delete students from the list (more free to add, interestingly, than to delete). In choosing modes of instruction, they exercised their judgment to accept, reject, or modify instructions that accompany packaged materials. And they valued highly opportunities to use their judgment in such matters as hiring aides, choosing texts and tests, allocating time, and choosing colleagues with whom to cooperate. Indeed, many of the examples of frustration or discontent that were mentioned involved perceived infringements on their right to exercise professional judgment.

Finally, in connection with the general topic of individualization, it is interesting to note that at least two of our authors (Amarel and Hawkins) urge caution in evaluating methods and modes where a faulty "more the better" hypothesis may be operating implicitly. Hawkins uses a medical analogy to describe a situation in which a needed treatment, overdone, becomes a destructive causal agent, and Amarel emphasizes the possibility of curvilinear relationships as the realities behind assumed linear descriptions. It is entirely possible, they suggest, that a particular instructional strategy might be highly effective used for, say, 10-minute periods and demonstrably destructive used for 30-minute spans.

We are left with the impression that the scientific portion of the IDS has told us what not to count on for results in achievement, and this is helpful. The teachers, for their part, have told us what we may be able to count on, but we cannot yet transform this information into prescriptions. Perhaps the wisest course for a policymaker is not to try to prescribe exactly or narrowly what form instruction should take but to support responsible experimentation, continuous and dual evaluation of the sort attempted by NIE in the IDS, and openness in the dialogue between teachers and researchers.

Summary

As the IDS itself yielded important information about teaching as science, the IDS Conference filled out that picture with a glimpse of teaching as art and profession. Researchers approach teaching situations indirectly and analytically through the mediation of variables and hypotheses. Teachers approach teaching directly and intuitively. It has long been recognized that researchers can contribute information useful for teachers. Now we see that teachers can make a contribution toward more significant research.

Their discussion suggests that it might be worth concentrated research effort to get more information in the following areas:

(1) Curriculum materials--How teachers use, adapt, and refine them; whether teachers understand their potentials; how favored instructional modes influence the choice of materials

(2) Instructional modes--How these are matched to student needs and curriculum patterns; how many distinct modes a successful teacher uses effectively; under what conditions various modes are effective

(3) Teacher flexibility--How teachers adapt to varying needs in their students; use of various modes in instructional discourse; adaptation and refinement of objectives as well as materials

(4) Enhancement of self-image as an educational goal--Can it be shown that children really do feel better about themselves--and with some reason--when they have worked with teachers who place self-image high on their list of educational goals? What do teachers do to enhance self-image? How often are the techniques that conference teachers reported using actually used by other teachers?

Teachers and researchers both profit from the sort of dialogue initiated by NIE. But policymakers also profit. The information teachers contribute tends to soften research results and to caution policymakers about simple solutions. The conference teachers urge us to look beneath and beyond quantitative findings for the qualities that make a difference in effective teaching. A

conference of this sort points up, rather dramatically, that there is no "royal road" to the understanding of educational problems. Rather, high orders of tolerance for ambiguity and of critical thinking skills are required of teachers, researchers, and policymakers alike.

Chapter III. Educational Opportunity

OPPORTUNITY TO LEARN

Gaea Leinhardt
Learning Research and Development Center

The charge for this paper was to integrate the comments of teachers with the research literature in the area of opportunity to learn. Of the 700 teachers who participated in the Instructional Dimensions Study (IDS) (Kirschner Associates, Inc., 1977), 40 gathered at a conference to discuss the results of that study. This paper attempts to weave the ideas and concerns voiced by 40 teachers and specialists over a 2-day period with the ideas and concerns voiced by researchers over the last 10 to 15 years. If there appears to be an imbalance in the presentation, it is due in part to an imbalance in the availability and form of the information. The paper is not a complete review of the literature in the area. Rather, I have tried to treat the major themes extant in the literature as a backdrop for the dialogue between the study findings and the teachers.

This review focuses on the concept of educational opportunity at the classroom level. Examination of educational opportunity at the system and school level has been the focus of a considerable amount of previous effort (Coleman et al., 1966; Jencks et al., 1972). Although the specific variables of concern are very different, the intent is similar. Opportunity is the chance of obtaining, or the access one has to, a valued resource. It is essentially an economic and political concept as opposed to a psychological or sociological one. In education, the resource of interest is knowledge. The more commonly described or measured resource is formal education, access to which presumably leads to knowledge.

The construct of opportunity can cover a wide range of concepts, from physical facilities to faculty credentials to student learning behaviors. How opportunity is measured in any one situation tends to define it. In the IDS, opportunity was one of four instructional constructs thought to affect learning outcomes; the other three were motivators, instructional events, and curricular structure (Cooley and Leinhardt, 1975a, 1975b, 1978). Opportunity was operationally defined by two variables: the amount of time available for reading and math instruction, and the impact of exposure to criterion relevant instruction.

Time

It is important to remember that the passage of time itself does not affect learning directly. It is only the time spent in doing something that is relevant. Time is a convenient metric of sequence and duration. The degree of convenience varies with the perspective of the viewer. To the degree that we can easily express parameters of instruction and learning by it, it is a very useful tool. It is the power of time to inform with respect to sequence that permits us to infer cause and effect.

In the IDS, time was measured in two different ways: (a) the quantity, in minutes, of supplemental instruction given to a student; and (b) the combination of: the number of days between fall and spring testing, the time scheduled for reading and math, attendance rates, and the percentage of students on-task, while large class size and a high turnover of students were considered to be detractors from time. When it is reported that time was positively related to achievement, it means that both of these basic representations of time were positively related to achievement.

Researchers in education have found it useful when describing the time available to students to distinguish between different levels of time measurement. When describing the time available for subject matter, researchers tend to distinguish between allocated and engaged time and between the amount of time given as opposed to the amount needed (Anderson, 1973; Berliner, 1976, 1977; Berliner and Rosenshine, 1976; Carroll, 1963).

Allocated time refers to the assignment of blocks of time to a given subject matter area at the district, school, or grade level; it is not a reflection of how much time a student needs or spends doing the subject. In general, allocated time is considered to be the upper boundary for engaged time.

Engaged time refers to the amount of time a teacher and the class actually spend in the subject area--allocated time minus management instructions and disruptions. Engaged time can be further refined to include only the amount of time a student spends working on "appropriate" learning material (Berliner, 1976; Arlin and Roth, 1978).

Before we go further, lest it sound as if researchers are systematic and precise with respect to the issues of time, it should be noted that the research definitions tend to slide back and forth between one type and another with great ease. Researchers tend to measure time in the most economical way they can while still answering the question of interest. Thus, Wiley and Harnischfeger (1975) and Husen (1972) began with such broad measures as the number of days of attendance and have moved to estimates of actual instructional time (Harnischfeger and Wiley, 1977). Bloom and his students have tended to measure time on-task and found positive relationships where allocation and curriculum content were fixed (Bloom, 1976; Block, 1970).

Many large-scale studies, the IDS included, focus primarily on allocated time (Stallings, 1975; Beginning Teacher Evaluation Study [BTES], 1976-77; Kirschner Associates, Inc., 1977) because it is very costly and difficult to get more refined estimates. Often, as in the IDS, there is an attempt to modify the simple estimate of allocated time by taking into account the attendance of children, time on-task, etc. Studies investigating the relationship between allocated time only and achievement have rarely shown significant relationships between the two (Rosenshine, 1978; Stallings, 1975). For example, at the classroom level, very low correlations between minutes per week in a subject and gain (-.04 and .12) were found in the IDS. This is consistent with other work (Rosenshine, 1978). Wiley and Harnischfeger (1974) did find significant relationships between allocated time and achievement, but these were not replicated by Karweit (1976) in her reanalysis of the data--nor were they replicated by the IDS (Kirschner Associates, Inc., 1977).

At the IDS Conference, teachers consistently referred to allocated time as a fixed entity. If allocated time is indeed fixed at the school or district level, then any individual variation in teacher practices (with respect to the organization of time) would tend to be ignored by an estimate of time that considered only allocation.[1]

More fine-grained estimates of time spent in instruction tend to result in stronger relationships with achievement (Anderson, 1973; Bloom, 1976; Fisher et al., 1977; Felsenthal and Kirsch, 1978; Welch and Bridgham, 1968). These estimates, unlike allocated time, discriminate among those teachers who are efficient managers and planners and those who are less efficient. Thus, two teachers working under similar allocations of time can, and probably do, differ in the amount of time their students are engaged in learning. The line between engaged and allocated time is not always clear in actual research (Marliave et al., 1977; Filby et al., 1977; Wiley and Harnischfeger, 1975). As methods of describing allocation get more and more precise, for example, allocation of time to place value, or compound words (Marliave et al., 1977), allocated time begins to look, from a descriptive sense, more like engaged time. The major difference, however, lies with the student. Estimates of allocated time tend to be free of student information. Estimates of engaged time tend to include student behaviors. As estimates get more precise and include student information, the relationship to achievement gets stronger.

In the descriptions of the specific activities students engage in during the language arts or reading period, it appears that considerable time is spent in nonreading behaviors. That is not to say that these activities are unimportant or not valuable, only that they are not reading. For most classes, the amount of time really spent in the subject area can probably be approximated by allocated time minus off-task times a fraction. This fraction has as a numerator the number of adults and as a denominator the number of groups. For example, assume 100 minutes of allocated time minus 10% off-task, leaving 90 minutes; assume five reading groups with two adults; the fraction is 2/5 times 90, or 36 minutes. The assumption is that much, if not all, teacher-led time is academically focused while the majority of the remainder is not academically focused. It may be the case that higher amounts of time can be recorded, but it isn't clear what the children are doing during those times. There is almost always a discrepancy between teachers and researchers when estimating engaged or on-task time. Teachers consistently report a higher estimate of engaged time (80-90%) while researchers report a lower estimate (50-70%) (Good and Beckerman, 1978; Powell and Cahan, 1977; BTES, 1976-77). This may be because of differing definitions or perspectives, or one group may be correct. The point is that when researchers estimate engaged time, that estimate is related to achievement. I have not found any study in which the teacher's estimate of engagement was used as a predictor.

---

[1] For a fuller discussion of the implications of alternative modes of education, see Kohlberg and Mayer (1972).

Another way to view the problem is to consider what part of subject matter time is clearly nonreading or nonmath. It seems from teacher descriptions that there was between 20 and 40 minutes of nonreading time per hour. That time is spent in a lot of things: cutting and pasting, group discussion under the rubric of language development, or some type of group activity, such as preparation for a newspaper or a play which has a strong reading or writing component in it, but in which the number of minutes children are really reading is very small.

At the IDS Conference, teachers identified another source of reading time, content areas. Some teachers reported that actual reading time could be increased or reinforced through the content areas, rather than replacing content areas with expanded reading periods. The tradeoff, then, is between reading taking up larger amounts of time and increasing the time spent in those subjects that have a high reading component. By increasing content area reading, it is possible to increase the amount of engaged reading time above the level of allocated reading time. From an economic perspective, the opportunity cost of increasing allocated reading time above 1½ hours a day seems to be quite high, but the opportunity cost of increasing engaged reading time is not as high if one is willing to count reading that goes on in subject matter areas. This would not be as effective in mathematics, although science and parts of social studies can certainly reinforce early arithmetic skills.

Teachers mentioned some factors that tend to reduce instructional time such as: testing, individualization, physical movement, organizing large numbers of students, program or activity management, and discipline. In some districts, the formal testing load was incredible, covering as much as 3 weeks in the fall and spring. This can be reduced by coordinating Federal, state, and local testing requirements and keeping them to a minimum. Individualization, or independent work by children at different levels of study, was seen as reducing on-task behavior. It was clear that managing children who were working independently created some difficulty. One presumes that with an appropriate management procedure in place, this could be overcome (Wang, 1976). Moving students from one part of a building to another takes considerable time, but often is physically and economically impossible to avoid. It appears that movement is facilitated when adults accompany children and when the routines are well established and consistently implemented. The general area of discipline and program management came up as a problem mainly in cases where: (a) the supplemental programs involved handling large numbers of students simultaneously; (b) the program arrangements were being changed frequently; and (c) the links between regular teachers and supplemental teachers were not formally maintained by such activities as planning a student's program. In general, within-class discipline was not mentioned by teachers (although it frequently is in the popular press) as a major problem or source of time away from instruction.

An alternative way to increase "productivity" is to increase the rate, or pace, with which material is covered within a given period of time. Researchers have generally assumed that pacing is a function of student needs (Anderson, 1973; Bloom, 1974) and that smarter children are faster. However, recent research and common sense indicate high within-pupil variation on time to criterion (Glasnapp et al., 1978). A student may be fast at some things and slow at others within a single academic area, such as math. In general, teachers did not feel it was very desirable to step up the pace of learning. Several teachers felt that increasing pace would decrease mastery; however, this has not been supported by research literature (Barr, 1973-74). Teachers seemed more concerned with finding the appropriate pace rather than in speeding it up. The question of who should decide the optimum pace was discussed by the teachers. In general, teachers feel that they should set the pace of instruction. Most teachers seemed to feel that students would not pace themselves appropriately if left totally alone. In individualized or partially individualized programs where students share pacing responsibility with teachers, it was felt that teacher guidance was definitely needed. Dahllof (1971) has suggested that teachers tend to set level and pace at the bottom 25% of the class. When this is considered with Barr's (1973-74) finding that slowing the pace for the poorer children did not help them and speeding it up didn't hurt them, it may be that teachers, too, need some guidance on pacing.

Timing. In addition to learning about how much time is spent in a given area, researchers and teachers are interested in the arrangement of time, or timing (Karweit, 1977). Do students learn more from a solid block of instruction or more from several shorter blocks (Karweit, 1977; Rosenshine, 1978)? Is it important when instruction occurs within the year, the week, or the day? In general, the literature tends to reflect the view that continuous instruction is "better."

Continuous instruction means uninterrupted blocks of time within a day, daily presentation of material, and continuous 8- or 9-month school years. These features minimize tool-up and tool-down times. But as Karweit (1977) points out, there is a tradeoff between the tool-up costs and the costs of boredom. Some researchers feel that

all interruptions are to be minimized and that, given the very low estimates of time spent in subject areas, there is considerable leeway before students get bored (Rosenshine, 1978). Little distinction is drawn between planned interruptions and unexpected disruptions, although the latter is probably more detrimental than the former (Kounin and Gump, 1974; Kounin and Doyle, 1975; Karweit, 1977). Teachers seemed to have a more subtle conception of arrangement of time. They described a base core of uninterrupted time to which additional time could be added around it (long/short/short or short/long/short). One teacher reported using two large blocks of time--a.m. and p.m. (long/long)--in which different instructional approaches for the same content were used (linguistic in the a.m. and whole word in the p.m.).

Another block of time to be considered is the school year and the possibility of rearranging it. The current arrangement consists of approximately 36 weeks of school with two or three major breaks. Other arrangements are clearly possible. The advantages to alternative arrangements seem to be less educational than social and economical. Some teachers were interested in considering arrangements which took the summer vacation block and distributed it throughout the school year, but most teachers felt that the tool-up time required at the end of a long holiday was such that you wouldn't want to do that more than once, and that children need a sustained break just as they need sustained instruction.

In summary, time is an important construct for research on education. However, we still know surprisingly little about it. Teachers and researchers need to develop a common and precise vocabulary that reflects quite closely the instructional content. Terms such as "language arts" are intentionally global and vague, and while they have the feature of integrating the many attributes of language usage, they tend to mask for both teachers and researchers how time is spent in the subcomponents. If labeling becomes more precise, then the measurement of time can follow. (This should not, however, be interpreted as support for a highly fragmented analytic approach to the teaching of reading.) In formulating recommendations, researchers must be very sensitive to the opportunity costs for teachers and students that are associated with increasing allocations of time in one or another areas. The opportunity costs in time clearly are not linear. Thus, the first 10-minute increment of math time costs less than the seventh increment (see Karweit, 1977; Walberg, 1977). In considering researchers' recommendations, teachers need to be sensitive to their own limitations in estimating student time in learning activities, as opposed to teacher time spent teaching (Marliave et al., 1977).

Activity Structures. Activity structures are a way of examining the life of a student in a schoolday (Kounin and Gump, 1974; Kounin and Doyle, 1975; Bossert, 1978; Westbury, 1978). It is a relatively old and recently renewed sociological concept. Activity structures examine the experience sequences of students (or other actors) throughout the day and label blocks of contiguous time not by the cognitive information, but by the activities or actions in which the child is actually engaged.

There are two major contributions that the concept of activity structures makes. First, it points up the vast variety of actions in which the actors in a school engage. Individuals who are not frequently inside the schools have only the dimmest of notions of what goes on in them. Second, the concept aids in keeping straight the many perspectives of the actors in a school. Time spent in mathematics has very different meanings when a researcher clocks a student or a teacher. In fact, that very distinction may resolve much of the problem associated with estimating time. Activity structures serve a useful descriptive and definitional function.

The notion of activity structures must be kept in perspective, however. As Doyle (1978) points out, the notion of activity structures is useful and fascinating if what one wants to do is describe internal relationships of elements of schooling. Doyle also points out that activities are very important to teachers and form a core for planning. If, however, one wants to draw causal relationships between instructional procedures and student learning, then the notion of a task is probably more useful. Tasks can be viewed as having goals and mechanisms for meeting the goals (Doyle, 1978). In the IDS, the object was to identify the dimensions of instruction that affect achievement, not to describe the total instructional environment. The limitation to the task approach is that important causal events may be missed by the researchers' definitions--the limitation of the activities approach is data swamp.

Overlap

The second major variable within the opportunity construct is overlap. Overlap refers to an estimate of the degree to which the curriculum teaches the material assessed by the criterion measure (test). Overlap is a part of the opportunity construct because it represents the chance the student had to be exposed to criterion-relevant material. Overlap does not estimate how much of what was taught was tested, only how much of what was tested was taught. It is important to include information on overlap in any study of instructional impact because students could have

a very "good" educational program and spend considerable time on it, but there might be little criterion-relevant (test-relevant) instruction in the program.

An estimate of overlap can be used in three ways: (a) it can be used to "adjust" the criterion measure by deleting all items not taught in the curriculum; (b) it can be used to stratify classrooms or only consider comparisons of instructional technique among classrooms with the same (or similar) level of overlap; and (c) it can be used as one of the variables in an analysis such as regression. In the IDS, the third option was chosen because options (a) and (b) made it impossible to compare programs simultaneously.

The question of how to measure overlap is somewhat complex. As with time, the specific procedure for measuring overlap will affect how one should interpret the results. There are two basic approaches to estimating overlap. The first is to label either the subtests or subcategories of the test(s) and estimate both the presence and emphasis of these categories in the curriculum. In some cases, this is done by including information about student end-of-year location (Kugle and Calkins, 1976); in others, it is the overall curriculum that is compared (Armbruster et al., 1977). When general overlap estimates are correlated with other curricula and achievement gains, a significant positive relationship is found for reading comprehension (Armbruster et al., 1977). Thus, to the degree that a reading text emphasizes the same areas of reading as the test, the students will perform better.

The second approach is to define what is needed in order to pass each item on the criterion measure, then to search the curriculum to see if and where the information is taught. This type of analysis is done at the student level (Cooley and Leinhardt, 1975a, 1978; Poynor, 1977a, 1977b), and it is more fine-grained and precise than other approaches to measuring opportunity.

The IDS finding was that there was a high positive relationship between overlap and test performance. This finding is especially important because of its conservative nature. It was assumed in constructing the measure that children had the capability of generalizing and that that would reduce the power of the overlap measure. That is, children would be able to get items correct to which they had not necessarily been directly exposed. While that is undoubtedly true, the probability that the student will get the items correct seems to be substantially higher if the student has been exposed to the specific material in that item, and that is what's important.

In the IDS, overlap was measured by creating a dictionary of the Comprehensive Tests of Basic Skills (CTBS) (CTB/McGraw-Hill, 1973-74). Then a dictionary of all the major curricula being used in all of the regular and supplemental classrooms was made (Poynor, 1977a). The location of each student at the end of the year in that curriculum was monitored, and the cumulation of all the curricular information was mapped onto the specific test information. For each item on the CTBS, it was necessary to determine if the student had been exposed to material that would permit him or her to get the item correct. Specific rules were generated. For example, if an addition problem was encountered on the CTBS, then the student must have had either those specific numerals or numerals higher in value than those numerals. Thus, if a student had been exposed to 8 and 9 in addition problems, and the problem was 6 and 5, then the overlap was a positive one. For the reading items, an analysis was done to determine which words a student needed to know in order to answer the test items. The number of words was generally less than the number of words that were in the specific item, and both the stimulus and response words had to have been taught at some place in the curriculum. The massiveness of this task can only be appreciated when one realizes that each student is actually exposed to multiple curricula, quite frequently in each of the subject matter areas, reading and math.

When presented with the IDS findings, teachers agreed that children would do better when tested on material that they knew and had been taught. However, the teachers tended to expect students to be able to perform on tests that require the student to generalize to a greater extent than the IDS finding implies. Teachers felt that item form was quite important in predicting how easily a student could respond to a given problem. For example, if a student had never been exposed to horizontal addition, it was unlikely the student would read the question correctly and respond correctly, even if he or she had been exposed to vertical addition. In general, they reported teaching children the format, although not the content, of testing prior to the test.

Testing. Teachers used the discussion of overlap to go into considerable detail about their attitudes toward testing. The discussion revolved around three aspects of testing: (a) teachers' attitudes toward standardized, nationally normed tests, (b) teachers' attitudes toward nonstandardized criterion- or domain-referenced tests, and (c) teachers' views of a national testing program.

Teachers responded favorably to standardized testing programs if they (or their representatives) had been involved in selecting the test. It was especially important that the test reflect the major curricular objectives of the statewide curriculum being used. Teachers agreed, when

pressed, that testing gave them useful information. For example they felt tests provided an accurate assessment of who should be eligible for Title I services. For the most part, the tests separated the top-performing children from the low-performing children. Teachers did not object to the norming process itself, and they felt it was reasonable to know how their children were doing in relationship to other children in the State, district, and country.

However, tests can be and frequently are abused. Often there were serious problems reported with overtesting. There were some cases where tests seemed to be inappropriately used to judge programs and children. Teachers felt that standardized testing is difficult for children in their early years, and it is specifically difficult for Title I-eligible compensatory education students. Teachers also felt that no standardized test ever gave them information at the student level that could be used for teaching purposes.

A major alternative to standardized testing that is being used in some school districts is criterion- or domain-referenced tests. Teachers described several scenarios in which criterion-based tests were effectively used. One mode is to test students according to a time schedule, minimally at the end of every year, maximally every 3 or 4 months. Another approach is to request a test whenever a student or a group of students is ready for a specific concept to be tested. Some districts have developed a testing bank for individual objectives that had been agreed upon by both the teachers and the district-level personnel. Criterion-based testing was perceived as being just as bad as standardized testing if it was externally laid on. It is only superior in the eyes of teachers when it is used with teacher input and remains flexible with the option of teachers revising it.

Criterion-based testing tends to be helpful in diagnosing individual student needs. This may be because the terminology used in the tests and the item forms are very similar to what the teachers are teaching. It seemed to be an assumed characteristic of criterion-based tests that they were manipulative as opposed to multiple choice.

Many researchers share teachers' concerns and misgivings about testing. Several elements seem to be relevant to this problem. Tests are very easy, especially standardized tests, and they tend to reflect only the superficial elements of knowledge. They are easy to give, to score, and, in some ways, at least, to interpret. Highly structured and focused instruction yields impressive gains on highly structured and focused tests. But questions on whether these techniques retain their impact in more complex examinations of knowledge remain unanswered because the more subtle and complex elements in knowledge acquisition remain unmeasured. While many teachers and researchers (Doyle, 1978) believe that tightly structured instruction will fail to be as important when broader knowledge is measured, their view, for now at least, remains speculative.

A final serendipitous point which emerged from the discussion was the concept of a national testing program. Researchers and politicians have tended to back away from the whole aspect of nationally based testing, especially in the area of basic skills. Teachers, however, were very open and more receptive to that concept than would have been expected. Teachers thought that a national testing program could be developed if it were done in strands (maybe of objectives), if representative teachers were deeply involved in the construction, and if the test was manipulative as opposed to multiple choice. They see considerable utility in having a national test, in part because it could help reduce the overall testing load.

Teachers were sensitive to the concept of appropriate criteria. This came out of discussions of some findings that were somewhat inconsistent with what the teachers expected. They pointed out that researchers tend to measure things that are easily measurable. For example, they felt that it was not reasonable to dismiss the concept of class size as being unimportant if no impact on achievement was found because the appropriate criterion for the impact of class size may not be achievement, but might be student and teacher emotional state or quality of life. However, they did admit that it would be far more powerful if something like class size could be shown to have detrimental effects on achievement.

Conclusions

The opportunity to learn is a valuable way of considering educational resources and is a useful heuristic for studying effective instructional features. It is important, however, to realize that opportunity is only one dimension and that other areas, such as quality of instruction, are also worth investigating. Research results indicate that when opportunity is carefully defined in terms of time spent working on criterion-relevant material or criterion overlap, then there is a strong relationship between it and achievement.

Opportunity should probably be limited to mean the quantity of time available or used and quantity of criterion-relevant instruction. While considerable research has been done in the area of time spent in specific learning activities, more precise knowledge is needed about the cost and differential impact of spending time in different subjects at different grade levels. We need to

know, for example, if 10% increments in math time will show more influence at 2d grade than 5th grade, or more or less than an equivalent increase in reading. Teachers and researchers seem to believe intuitively in a point of diminishing returns for increases in time, but where is it? Most people seem to believe that the significance of curriculum overlap diminishes as the criterion gets more complex and as children get older---but where and when does this occur? We will need both clinical and research-based information to answer these questions.

BIBLIOGRAPHY

Anderson, L. W. "Time and School Learning." Unpublished doctoral dissertation, University of Chicago, August 1973.

Arlin, M., and G. Roth. "Pupils' Use of Time While Reading Comics and Books." American Educational Research Journal 15(2):201-216, 1978.

Armbruster, B. B., R. J. Stevens, and B. Rosenshine. Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests. Technical Report 0.26. Urbana, Illinois: University of Illinois at Urbana-Champaign, 1977.

Barr, R. C. "Instructional Pace Differences and Their Effect on Reading Acquisition." Reading Research Quarterly 11(4):526-554, 1973-74.

Beginning Teacher Evaluation Study. Technical Report Series and Technical Note Series. San Francisco: Far West Laboratory for Educational Research and Development, 1976-77.

Berliner, D. C. "Instructional Time Allocation in Fifth Grade Reading." Beginning Teacher Evaluation Study, Technical Report II-5. San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Berliner, D. C. "Instructional Time in Research on Teaching." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Berliner, D., and B. Rosenshine. "The Acquisition of Knowledge in the Classroom." Beginning Teacher Evaluation Study, Technical Report IV-1. San Francisco: Far West Laboratory for Educational Research and Development, 1976.

Block, J. H. "The Effects of Various Levels of Performance on Selected Cognitive, Affective, and Time Variables." Unpublished doctoral dissertation, University of Chicago, 1970.

Bloom, B. S. "Time and Learning." American Psychologist 29:682-688, 1974.

Bloom, B. S. Human Characteristics and School Learning. New York: McGraw-Hill, 1976.

Bossert, S. T. "Activity Structures and Student Outcomes." Unpublished manuscript, University of Michigan. Paper prepared for the NIE National Invitational Conference on School Organization and Effects, San Diego, January 1978.

Carroll, J. B. "A Model of School Learning." Teachers College Record 64:723-733, 1963.

Coleman, J. S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfeld, and R. L. York. Equality of Educational Opportunity. Washington, D.C.: U.S. Government Printing Office, 1966.

Cooley, W. W., and G. Leinhardt. The Application of a Model for Investigating Classroom Processes. Publication 1975/24. Pittsburgh, Pennsylvania: University of Pittsburgh, Learning Research and Development Center, 1975a.

Cooley, W. W., and G. Leinhardt. Design for the Individualized Instruction Study: A Study of the Effectiveness of Individualized Instruction in the Teaching of Reading and Mathematics in Compensatory Education Programs. Pittsburgh, Pennsylvania: University of Pittsburgh, Learning Research and Development Center, 1975b.

Cooley, W. W., and G. Leinhardt. IDS: The Search for Effective Classroom Processes. Final report submitted to the National Institute of Education, 1978.

CTB/McGraw-Hill. Comprehensive Tests of Basic Skills, Level I, Form S and T; Level B, Form S. Monterey, California: McGraw-Hill, Inc., 1973-74.

Dahllof, U. S. Ability Grouping, Content Validity, and Curriculum Process Analysis. New York: Teachers College Press, 1971.

Doyle, W. "Student Mediating Responses in Teaching Effectiveness: An Interim Report." Paper presented at the Annual

Meeting of the American Educational Research Association, Toronto, March 1978.

Felsenthal, H., and I. Kirsch. "Variations in Teachers' Management of and Time Spent on Reading Instruction: Effects on Student Learning." Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Filby, N. N., R. S. Marliave, and C. W. Fisher. "Allocated and Engaged Time in Different Content Areas of Second and Fifth Grade Reading and Mathematics Curriculum." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Fisher, C. W., N. N. Filby, and R. S. Marliave. "Instructional Time and Student Achievement in Second Grade Reading and Mathematics." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Glasnapp, D. R., W. L. Deaton, and H. Sokolove. "Aptitude Achievement and Time to Criterion." Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Good, T. L., and T. M. Beckerman. "Time on Task: A Naturalistic Study in Sixth Grade Classrooms." Elementary School Journal 78(3):193-201, 1978.

Harnischfeger, A., and D. E. Wiley. "Time Allocations in Fifth Grade Reading." Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Harnischfeger, A., and D. E. Wiley. Teacher Resource Allocation: Consequences for Pupils. Chicago, Illinois: CEMREL, ML-Group for Policy Studies in Education, 1978.

Husen, T. "Does More Time in School Make a Difference?" Saturday Review, April 29, 1972, pp. 32-35.

Jencks, C., M. Smith, H. Acland, M. J. Bane, D. Cohen, H. Gintis, G. Heyns, and S. Michelson. Inequality: A Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, 1972.

Karweit, N. "A Reanalysis of the Effect of Quantity of Schooling on Achievement." Sociology of Education 49(3):236-246, 1976.

Karweit, N. "The Organization of Time in Schools: Time Scales and Learning." Unpublished manuscript, Center for Social Organization of Schools, Johns Hopkins University, Baltimore, Maryland, 1977.

Kirschner Associates, Inc. Instructional Dimensions Study. Final report. Washington, D.C.: Kirschner Associates, Inc., 1977.

Kounin, J. S., and P. H. Doyle. "Degree of Continuity of a Lesson's Signal System and Task Involvement of Children." Journal of Educational Psychology 67:159-164, 1975.

Kounin, J. S., and P. Gump. "Signal Systems of Lesson Settings and the Task-Related Behavior of Preschool Children." Journal of Educational Psychology 66(4):554-562, 1974.

Kugle, C. L., and S. Calkins. The Effect of Considering Student Opportunity to Learn in Teacher Behavior Research. Research Report No. 7. The Evaluation of Teaching Project. Austin, Texas: R&D Center for Teacher Education, University of Texas at Austin, 1976.

Marliave, R., C. W. Fisher, and N. N. Filby. "Alternative Procedures for Collecting Instructional Time Data: When Can You Ask the Teacher and When Must You Observe for Yourself?" Paper presented at the Annual Meeting of the American Educational Research Association, New York, April 1977.

Powell, M., and L. S. Cahan. "Policy Issues Related to Research on Instructional Time." Beginning Teacher Evaluation Study. San Francisco: Far West Laboratory for Educational Research and Development, 1977.

Poynor, L. H. Curriculum Analysis Procedures. Supporting report of the Instructional Dimensions Study. Washington, D.C.: Kirschner Associates, Inc., 1977a.

Poynor, L. H. Description and Measurement of Compensatory Programs. Santa Monica, California: System Development Corp., 1977b.

Rosenshine, B. V. "Academic Engaged Time, Content Covered, and Direct Instruction." Unpublished manuscript, University of Illinois at Urbana-Champaign, 1978.

Stallings, J. "Implementation and Child Effects of Teaching Practices in Follow Through Classrooms." Monographs of the Society for Research in Child Development 40(7-8, Serial No. 163):1-119, 1975.

Walberg, H. J. "A Theory of Educational Productivity." Unpublished manuscript, Institute for Research on Teaching, Michigan State University and the University of Illinois at Chicago Circle, 1977.

Wang, M. C. (ed.). The Self-Schedule System for Instructional-Learning Management in Adaptive School Learning Environments. Publication 1976/9. Pittsburgh, Pennsylvania: University of Pittsburgh, Learning Research and Development Center, 1976.

Welch, W. W., and R. G. Bridgham. "Physics Achievement Gains as a Function of Teaching Duration." School Science and Mathematics 68:449-454, 1968.

Westbury, I. "Activity Structures and Pupil Outcomes." Unpublished manuscript, University of Illinois at Urbana-Champaign, 1978.

Wiley, D. E., and A. Harnischfeger. "Explosion of a Myth: Quantity of Schooling and Exposure to Instruction, Major Educational Vehicles." Educational Researcher 33(4):7-12, 1974.

Wiley, D. E., and A. Harnischfeger. "Distinct Pupils, Distinctive Schooling: Individual Differences in Exposure to Instructional Activities." Paper presented at the Annual Meeting of the American Educational Research Association, Washington, D.C., 1975.

OPPORTUNITY TO LEARN: ANOTHER CASE OF THE RESEARCH TALE WAGGING THE DOG

Herbert Zimiles
Bank Street College of Education

At first glance, the new research interest in "opportunity to learn" and "engaged time" would seem to herald an advance in the methodology and conceptualization of educational evaluation. The results of evaluation studies, such as the IDS, can be better understood if we have detailed information about the learning environment whose impact is being assessed. In this regard, the need to describe and measure opportunity to learn--the extent to which a child has actually been available and/or receptive to the influence of his learning environment--is fundamental. However, the ambiguous nature of the definitions and interpretations of these new constructs raises questions about their usefulness. This paper examines some of these problems of definition and data interpretation, and also considers the circularity and self-fulfilling quality of research thus far stimulated.

The Problems of Definition and Measurement

In measuring opportunity to learn and engaged time, the educational researcher attempts to gauge the degree to which a child has actually had the opportunity to be influenced by the program whose impact is being evaluated. Such data can help to illuminate evaluation findings. If posttest results fail to show evidence of achievement test gains, how are these negative findings to be explained? Are the results attributable to a failure to implement the program properly or to the program's intrinsic limitations? If positive results are obtained, which aspects are mainly responsible for its success? Without detailed information regarding the transactions that take place in an educational program, it is almost impossible to account for its success or failure.

Efforts to study opportunity to learn and engaged time are directed at understanding the role of one aspect of the antecedent conditions in an evaluation study--the degree to which children are "tuned in" or actively participating in the work life of the classroom. The basic design of studies of opportunity to learn and engaged time calls for examining the correlation between measures of these variables and gains in achievement test scores. Such correlation data indicate the degree to which measures of opportunity to learn are predictive of educational achievement.

In pursuing this line of research, investigators tend to use different measures of opportunity to learn interchangeably. If it is established that a child (a) was absent much of the time, or (b) was seldom actively engaged in the work that was assigned, or (c) was not exposed to the material that was contained in the posttest measure of achievement, we may be helped to account for some of the results of an evaluation study. The first example focuses on whether or not the child actually attends class. The second deals with the degree to which the child is psychologically related to the events of the classroom and the learning situation. The third reflects a more functional analysis of the problem: instead of dealing with the child's pattern of reaction to the classroom, it asks whether he was exposed to the material included in the achievement test, whether he had an opportunity to learn what was being used as an index of progress. Two of the antecedents are concerned with characteristics of the student, with his physical and psychological presence; the third deals with characteristics of the teaching situation, with what the child has been exposed to in class.

The first, physical presence, would appear to be a necessary condition for school influence but hardly a sufficient one. It is an attribute that is easily and precisely measured. A child's "engagedness" in the classroom is much more difficult to define and measure. Engaged time is sometimes measured directly by observation and monitoring of classroom interaction and sometimes retrospectively by tallying the number of pages read or workbook assignments completed. Both kinds of indices present problems. Those based on direct observation of the child must define and measure engagedness in children, and definitions may differ. To the traditional educator, a child who is block building is not engaged in learning. Others may question whether the child who is ritualistically copying words or numbers from a workbook is actually engaged in learning activities. As to those indices of engagedness

that are based on the child's previous output, they may reflect ability level reather than engagedness.

Efforts to determine the degree of correspondence between exposure to particular content and performance on achievement tests present both conceptual and methodological problems. The correlation between posttest scores and exposure to test-related content in the classroom reveals more about the strengths and weaknesses of achievement testing than it does about the influence of opportunity to learn on achievement gains. In addition to being difficult to interpret, such data are marred by the procedures used to measure exposure to test-relevant content--teachers' retrospective judgments of whether the test material had been previously covered.

Despite the fact that these various antecedent measures assess different factors, present varying limitations and deficiencies of measurement, and generate data that are difficult to interpret, the results of research in this area have been cited as though they are a homogeneous, uncomplicated mass that has unmistakable implications for educational policy (see Rosenshine, 1978).

The Methodological Premise: Commitment to Evaluation by Means of Achievement Testing

Research in opportunity to learn accepts the basic methodological premise of most evaluation studies--that achievement tests provide valid measures of educational progress. This new work obscures longstanding perceived deficiencies in the methodology of evaluation. Educational evaluation has been subjected to two major lines of criticism: (1) achievement tests are regarded as invalid instruments of educational evaluation because they are only marginally related to many educators' hierarchy of educational objectives, and (2) methods of educational evaluation fail to identify the elements of the educational program that have brought about the gains recorded by the evaluation. The first of these criticisms is more sweeping but less universally offered; the second is less controversial. In addressing the second criticism, studies of opportunity to learn are, in effect, bypassing the first. Yet, achievement tests have been summarily rejected as a valid basis for educational evaluation by large numbers of educators. Among the reasons given are:

(1) Achievement tests are restricted in their content to factual material and specific skills. Such tests fail to indicate whether or not the child has been grounded in the material to be learned, how it fits into his cognitive framework, whether he understands why and how the information is to be used.

(2) In dealing exclusively with factual information and concrete skills, achievement tests fail to assess the more generalizable skills of cognitive functioning--problem-solving ability, originality and resourcefulness, and the ability to communicate ideas and information.

(3) Achievement tests do not deal with the affective dimensions of school experience--with the child's feelings about himself, his degree of self-knowledge, and his ability to relate to others.

These criticisms maintain that achievement tests provide a skewed assessment of how children are influenced by their school experience. Moreover, the measurement of those aspects that are covered by the test is flawed by the speeded, paper-and-pencil, multiple-choice format of the test. The test format dictates an undue emphasis on discrete fragments of information. As a result, test performance of young children is largely determined by the rate at which they have learned to read and by the amount of experience they have had in dealing with worksheets. Research in opportunity to learn and engaged time gives the appearance that these familiar criticisms of achievement tests do not exist. Without explicitly endorsing achievement tests, research in opportunity to learn and engaged time is built on a foundation of achievement test data.

Opportunity to Learn and Engaged Time as Intervening Variables Rather Than Causal Factors

While the variables of opportunity to learn and engaged time are clearly antecedent to the measurement of achievement gains, they are at the same time the consequences of the child's ability to cope with the expectations and demands of school life. The depth of a child's learning is influenced by characteristics of the child as well as by the quality of instruction. Opportunity to learn and engaged time frequently mediate achievement gains; they do not necessarily bring them about. Some children fall behind in their school work because they have been absent, and others are absent because they have fallen behind (and then they fall further behind). Engaged time, too, is better viewed as resulting from an interaction between the characteristics of the learner and the mode of teaching than as a manipulable determinant of educational progress. How much time a child spends in an activity in school is a reflection of many factors. If Dutch educational psychologists had conducted a study of Leiden elementary schoolchildren in the late 1600's, they would have been wrong to conclude that, if all other 2d graders had spent as much time drawing and painting as little Rembrandt van Rijns (more than three times the median for his age level), they, too, would have developed a fine proficiency

in the graphic arts. On the other hand, observations in the Leipzig schools 40 years later would have yielded other results. Gottfried von Leibniz was so swift in calculating that he was in the lowest decile of engaged time in mathematics seatwork. While most of his fellow 2nd graders were busily working out three-digit addition problems, Gottfried had long completed his assignments and was usually found reading a comic book. Engaged time scores do not always reflect teaching effectiveness, nor are they invariably associated with excellence of performance.

One way of sharpening the meaning of engaged time data is to contrast the variance of such scores obtained within and between classrooms. The within-classroom variance is more likely to be associated with individual differences in ability and responsiveness to classroom activity, whereas variance among different classrooms is likely to reflect variation in teaching styles. However, there are pitfalls to this method of analysis. In some cases, variation in engaged time within the classroom may result from differences among work groups that reflect the teacher's organizational pattern, her teaching style. In other instances, a comparison among classes that are comprised of different levels of homogeneous grouping or that were selected from schools with markedly different populations could spuriously inflate the between-schools variance.

The interpretation of engaged time data is further clouded by the fact that teachers do not simply have children randomly assigned to them. Teachers vary markedly in their capacity for tolerating the child with learning difficulties; those who use methods that are more supportive of children with learning problems are more likely to be assigned such children. Correspondingly, teachers who adopt high-pressure didactic methods often arrange transfers for those children who do not fit in with their demanding pattern of instruction. The study of engaged time will surely be influenced by the extent to which teachers welcome and support slow learners. Engaged time data may be expected to be consistently lower in classrooms of teachers who receive large numbers of slow learners because they are willing to work with them. This factor complicates our interpretation of correlations between engaged time and teaching style.

Misleading Aspects of a Common Thread of Theory, Method, and Research Findings

The new research on opportunity to learn and engaged time illustrates the manner in which the study of a problem is co-opted by key methodological decisions. One decision dictates a second that, in turn, leads to a third. They result in a series of interconnected events, mutually reinforcing and circular, that lead to the formulation of a study in terms that predetermine the outcome. Once the study of opportunity to learn became tied to a conception of educational progress defined by achievement test gains, the research became embedded in a theoretical framework of traditional education. If educational impact is defined in terms of the acquisition of specific skills and bits of information that are measured by standardized achievement tests, then opportunity to learn and engaged time must be defined, correspondingly, in terms of the conditions that facilitate rote learning. What is curious about this pattern is that what might seem to be the expected order of decisions is reversed, the choice of basic method leads to the adoption of a particular theoretical framework, which, in turn, dictates further methodological decisions. In this mutually reinforcing manner, an internally consistent but biased conceptual-methodological framework emerges that narrows the scope of the study and foreshadows its results.

Those who view school influence from the developmental-interaction framework (see Shapiro and Biber, 1972), a perspective that emphasizes the development of competence (White, 1959) and of coping skills (Murphy, 1962), are excluded by this perception of the problem. From the developmental-interaction viewpoint, schools function to strengthen the coping skills of children by broadening and deepening their experience and understanding of the world. Since each child's experience is distinctive, teaching and learning have to be individualized. The development of competence is concerned with fostering autonomy, resourcefulness and curiosity, problem-solving ability, and self-awareness. It is apparent that these goals call for a different set of educational strategies. Children must have an opportunity to function autonomously--to move about freely, to chat and discuss, to feel and smell and see as well as to listen--in order to deepen and integrate their understanding. The curriculum is designed to fit and enrich the child's own experiential background. In effect, children are expected to be engaged differently in school. Thus, how we define engaged time depends on our educational theory and our educational objectives.

As opportunity to learn and engaged time are currently being studied, it is assumed that children who are attending to the teacher's speech or busily involved in writing or reading activity at their desks are participating more fully and effectively than children who are rapt in thought or who are talking with their peers. The validity of these questionable assumptions is being tested exclusively by means of the criterion of

achievement test performance, which is primarily sensitive to and designed to measure rote learning.[2]

Although the results of research on opportunity to learn and engaged time are just beginning to accumulate, and although those that have been completed have used quite different methods to define and measure the variables in question, the general pattern of findings indicates a positive correlation between achievement test gains and measures of opportunity to learn or engaged time. The results are hailed by Rosenshine (1978) because they seem to him to identify antecedents that can be manipulated to produce increases in achievement test scores. Citing the results obtained by Stallings and Kaskowitz (1974), who found that time spent in reading and mathematics was correlated with achievement, Rosenshine calls for research that will assess the impact of stepping up the pace of engaged time and identify optimal dosage levels. He seems unready to entertain an alternative explanation of many of the findings: children who spend more time working are better students to begin with. They are not better because they work more, they work more because they are better. If we follow Rosenshine's thinking to its logical conclusion, then the data that show that children who attend school more regularly display greater gains in achievement should lead to large-scale hiring of attendance officers whose job it will be to improve the attendance of lagging students--as if attendance by itself will produce the desired effect.

Thus, a set of interrelated decisions has transformed an interesting research question into an exercise that yields a predictable body of findings. The choice of achievement test gains as an index of educational progress has led to a definition of engaged time and opportunity to learn in terms of classroom activities and events that are consistent with the kind of learning that is measured on achievement tests. In choosing to interpret the correlations found between engaged time and achievement test gains as causally connected rather than as concomitant variables produced by a common third factor, researchers have arrived at the conclusion with which many of them started--that direct instruction is the most efficient approach to education. Instead of asking how each 5-hour day of schooling can be shaped to form a coherent and stimulating learning environment that attempts to meet the principal developmental needs of each child, the results of research on opportunity to learn lead to recommendations for adding 3-minute increments in reading and arithmetic instruction to the regimen of daily instruction. Further, by its use of achievement tests as definitive indices of educational progress, its way of defining and measuring engaged time, and the uncritical interpretation of findings, research in opportunity to learn has become a powerful supporter and promoter of traditional education.

Because the issues of opportunity to learn and engaged time are intrinsically interesting and offer potential for improving the efficiency of teaching, the apportionment and use of time does merit study. However, a functional analysis of this problem is exceedingly difficult to achieve because each transaction in the classroom serves a multiplicity of purposes. The use of time can be assessed from an infinity of perspectives. Thus far, efforts to study this problem have been marred by the choice of mechanical, theoretically uninteresting bases for differentiating the use of time, premature concern with measuring the efficiency of use of time by invoking criteria of effectiveness whose validity is widely questioned, and drawing unwarranted inferences from the findings.

The Instructional Dimensions Study Conference

When NIE convened 40 teachers who had participated in the IDS and asked them to discuss the usefulness and relevance of data regarding opportunity to learn, the results were illuminating. Much of the initial discussion dealt with allocated time rather than engaged time. Many, but not all, of the teachers emphasized that allocation of instructional time was governed by local or state regulations, so that they were not free to alter patterns of instructional time allocation. Nevertheless, they pointed out that there was more flexibility than appeared on the surface because reading and arithmetic could be embedded in the teaching of social studies.

For the most part, the teachers seemed very caught up in the mission of Title I and dedicated

---

[2] Before we get further involved in the discussion of time, mention should be made of mathematics. Mathematics may have been equally represented within each of the four groups of teachers, but it was not equally discussed. Therefore, the majority of the discussion in this review focuses on reading instruction. Some of it may be true for mathematics as well. But a clear feel for use of time in math is still lacking. It appears to me in reviewing my notes that there is a block of stable time allocated to mathematics of between 40 and 50 minutes a day. It doesn't have the wide fluctuations of between 40 minutes to 2½ hours that reading does. It appears that the major constraint on mathematics is the amount of time that is subtracted for management purposes. Once the teacher starts teaching math, most children are doing mathematics.

to achieving its objectives, but many raised questions about some aspects of the programing. Some teachers expressed concern over the double dose of reading and arithmetic instruction that characterized their Title I program because it is being achieved at the expense of time for science and social studies. Is the neglect of these areas of instruction, in effect, sowing the seeds of failure in future years? The lack of exposure to concepts and information in science and social studies in the early grades, when combined with the meager experiential background of many of the Title I children, might lead to the creation of even more serious academic problems in later grades.

Other teachers spoke of the importance of saving time for discussion and creating opportunities for children to express themselves, thereby supporting the development of their self-image and their sense of individuality. Some teachers felt that the critical needs of Title I programs related to the children's feelings about themselves and their limited ability to communicate; they believe that the focus on academic training is misguided.

Discussion also centered on the idea that it was not only the amount but the quality of instructional time that mattered. Small-group instruction was reported to be much more effective than large-group teaching. On the other hand, in some cases the extra, more intense instruction especially provided for Title I children was offered by a separate tutorial staff that functioned independently from the regular classroom teacher. As a result, this group of children who were especially in need of strengthening their integrative functioning was being subjected to a disjointed and fragmented mode of instruction.

Problems of discipline also affected decisions regarding which children should be tutored. There is continuing pressure to remove disruptive children from the class. Indeed, many children who were unable to benefit from classroom instruction functioned more effectively in the less rigid situation of a learning laboratory.

In describing how instructional time was being used in the classroom, it became apparent that many teachers in the IDS classrooms worked at a stepped-up pace. They referred to the variety of curriculum packages that their district was using to accelerate the pace of learning. For the most part, they seemed to like these additional curriculum and text/workbook materials. However, they believed that Title I teachers were in a better position to make decisions about the allocation of instructional time than district administrators who were less familiar with the needs of the children and the realities of classroom life. They seemed ready to accept a prescription of time allocation but wanted it to be made locally.

When asked how instructional time could be deployed more efficiently, most members of this highly motivated and capable group of teachers believed that classroom time was being used efficiently. They contended that most instances of children sitting around and not working were likely to be found among the brighter children who had already completed their work and were waiting for the others to finish so that they could move on.

The fact that children performed better on items whose content had been included in classroom instruction did not cause surprise. No teacher reported having access in advance to the test and giving children instruction or practice on the actual items of the test, but many teachers indicated that it was customary to provide training on the type of inquiry used by the test to be administered. In some districts, children are given extensive experience with the item types that will appear on the test they are about to take.

Many of the teachers spoke of the great amount of testing that was going on in their school districts, so much that it was making inroads into instructional time. Teachers seemed to have mixed feelings about the role of testing. While many felt that testing had become excessive, most accepted the notion that testing represented the best method for assessing the educational progress made by the children. They were well attuned to the needs for testing and the ins and outs of preparing children for this method of measuring their achievement. Some teachers deplored the fact that the entire program evaluation rested on test performance, questioning their validity by observing that tests measure what can be measured rather than what should be measured. They much preferred locally created tests, those devised on the basis of a local formulation of educational objectives. They disliked nationally normed tests, tests that were less relevant to the actualities of their instructional program and the characteristics of the children they were testing.[3]

The concept of opportunity to learn did not seem to this group of experienced teachers to be one that might lead to a revision in their way of teaching. They are too familiar with the realities of day-to-day teaching and the complexities of

---

[3] One solution to the overtesting problem might be to do an expanded anchor test study. This would permit cross-instrumentation comparisons without actually testing children. The current norming information is too weak to really use.

classroom life to embrace a concept that is introduced by the research world in abstract and simplistic terms, especially a concept with which they are already grappling in more sophisticated terms. The opportunity to discuss the results of the IDS with a group of teachers helped to shed new light on the source of variation of some of the independent variables. At the same time, it helped to lay bare the disparity between the concerns of the classroom teacher and those of the researcher. It also demonstrated how the methodological premises of researchers, as well as their findings, can come to influence educational policy.

BIBLIOGRAPHY

Kohlberg, L., and R. Mayer. "Development as the Aim of Education." Harvard Educational Review, Vol. 42, November 1972.

Murphy, L. B. The Widening World of Childhood: Paths Toward Mastery. New York: Basic Books, 1962.

Rosenshine, B. V. "Academic Engaged Time, Content Covered, and Direct Instruction." Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Shapiro, E., and B. Biber. "The Education of Young Children: A Developmental-Interaction Approach." Teachers College Record 74(1):55-79, September 1972.

Stallings, J., and D. Kaskowitz. Follow Through Classroom Observation Evaluation 1972-73. Menlo Park, California: Stanford Research Institute, 1974.

White, R. W. "Motivation Reconsidered: The Concep' of Competence." Psychological Review 66:297-333, 1959.

Chapter IV. Setting for Instruction

SCHOOL SETTING AND LEARNING

G. Charlette Kennedy
Institute for Research on Teaching
East Lansing, Michigan

Introduction

The Instructional Dimensions Study (IDS) examined the effects of selected compensatory education program characteristics on student achievement. One feature of compensatory instruction which was investigated was relative effectiveness of pullout and mainstream instruction.[1] The research showed the mainstream setting to be significantly better for grade 1 reading and arithmetic and grade 3 reading. For grade 3 math there was no difference.

Given the prevalence of pullout Title I instruction, the study results indicated a need for careful review of setting effects upon student achievement, particularly that of 1st-grade students. The IDS Conference was convened to discuss study findings with a sample of participating teachers. As part of the agenda of the conference, discussions were held to identify features of mainstream and pullout instruction that accounted for high student gains. The sessions on setting addressed the following questions:

(1) How are Title I programs being implemented?

(2) What impact does the Title I program have on students and teachers?

(3) What are the advantages and disadvantages of mainstream and pullout instruction?

Since the instructional setting element was used only as a blocking variable in the IDS (Poynor et al., 1977), teachers' responses to these questions provided a contextual frame for interpreting study findings. The report that follows is a practitioner's synthesis of teacher responses to the above questions, with accompanying reflections. This insider/outsider perspective will provide some insight to the elements of teaching activities and classroom experiences that contribute to what children learn in schools.

Practitioner Responses

Implementation. Teacher participants at the IDS workshop described Title I instructional settings in terms of who, where, how, and, occasionally, when compensatory instruction was delivered. The criteria teachers employed to categorize programs as "mainstream" or "pullout" were based on their perceptions of personal roles and responsibilities associated with initiating, reinforcing, or reteaching a skill or concept. Mainstream setting was typically used to refer to compensatory instruction provided to Title I beneficiaries in regular classrooms. The regular classroom teacher assumed primary responsibility for instructional planning; directing and managing the activities of other certified or noncertified teaching personnel; and monitoring, evaluating, and reporting student progress.

Teachers discussed two types of mainstream settings. In one, the regular classroom teacher provided compensatory instruction for Title I children in the regular classroom, assisted by auxiliary personnel such as paraprofessionals and preservice interns. Auxiliary personnel were funded by Title I, although school district resources often augmented Title I monies. The regular classroom teacher assumed total responsibility for identifying children's learning needs, organizing and planning instructional events, and directing the activities of auxiliary personnel.

In the second type of mainstream setting, a reading or math specialist or other Title I-funded staff person came into the regular classroom to provide instruction for children who qualified for

[1] Pullout instruction is defined as supplemental instruction that is delivered outside the regular classroom. Mainstream instruction is supplemental instruction delivered within the regular classroom.

assistance under Title I guidelines. The children were either referred to the specialist by the classroom teacher or identified through a testing program; often both procedures were used. With this type of setting, the classroom teacher and the Title I specialist jointly planned for the instruction of Title I students, but the classroom teacher initiated the instructional events and the Title I specialist reinforced the instruction.

In pullout settings, compensatory instruction was typically provided to Title I students in a space designated as a lab or resource room. A reading or math specialist assumed primary responsibility for instructional planning, for recommending to the classroom teacher other activities in which the learner might appropriately be engaged, and for monitoring and evaluating student progress.

Teachers described two varieties of pullout instruction. In one variety, students were referred by the classroom teacher or were identified through testing procedures. The regular classroom teacher and the specialist conferred formally as well as informally to monitor each child's progress and to plan future instruction. The specialist, however, initiated instruction in a skill area, and the regular classroom teacher reinforced the instruction by providing classroom experiences and activities which gave the child the opportunity to apply the skills learned.

In the second type of pullout instruction, the classroom teacher identified instructional needs of Title I children and sent them to the laboratory or resource room for this instruction. This room, usually equipped with special materials as well as hardware such as tape recorders and cassette viewers, was staffed by Title I personnel who may or may not have been certified teachers. The classroom teacher conferred with Title I personnel to monitor and evaluate student progress.

In a hybrid variety described by one reading specialist, entire classrooms attended the reading lab daily. The regular classroom teacher taught non-Title I youngsters at the same time that the reading specialist taught Title I children. The regular classroom teacher and the reading specialist jointly planned the reading instructional events for both groups of children. Instructional assistance was provided by two aides--the classroom teacher's aide and the reading specialist's aide.

Impact of Mainstream and Pullout Settings on Students. Workshop facilitators focused on several claims assembled from conventional wisdom that address the impact of mainstream and pullout instruction for students and teachers. Workshop participants' comments suggest that teachers are primarily concerned about children--their social development (i.e., peer interactions), their self-image, and their motivation level. Teachers' comments also addressed the issue of stigmatizing children.

According to conventional wisdom, mainstream instruction stigmatizes compensatory education students by having extra instruction occur in the presence of peers. Pullout instruction is believed to stigmatize compensatory education students by making them easily identifiable. Teachers advocating mainstream instruction indicated that peer and parental pressure result when children leave the room for instruction. A teacher from a Florida school district clearly and emphatically reported, "Parents don't like it when their children have more than one teacher."

A Michigan teacher favored pullout instruction because of the opportunity it provided for children to confront and overcome their perceived inadequacy. She said, "Children know they don't know how to read. The pullout situation gives them a chance to do something about it with someone who understands their problem."

Most teachers working in pullout programs acknowledged that compensatory education students were stigmatized and that this created management problems. Its negative impact on Title I youngsters was reduced by practices such as open and frank classroom discussions concerning academic, physical, and cultural differences; ongoing counseling for all children to encourage empathy and improve classroom relationships among peers; and providing opportunities for Title I children to perform successfully in other school and classroom situations.

Most of the teachers attending the workshop came from districts in which the pullout setting was preferred for delivering compensatory reading and mathematics instruction. Regarding stigmatization, they observed that students are not labeled in building or classroom environments characterized by frequent movement and a lot of activity (for example, special programs for gifted and talented youngsters).

Fewer reports of stigmatization were associated with compensatory education programs employing "systems" instructional formats in which all children were simultaneously engaged in learning tasks in a common skills area (for example, comprehension, study skills, or analytical skills). Systems instructional formats are characterized by continuous groupings and regroupings according to test performance. This procedure encouraged flexible grouping arrange-

ments. More importantly, teachers believed that students' perceptions of the similarity of instructional content area among class members minimized Title I children's feelings of "being different" and classmates' perceptions of Title I children as "weird."

Teachers attributed reduced stigmatization to the minimization of differences in instructional content areas to which compensatory and noncompensatory education students were exposed. Within each content area student objectives were formulated, children were grouped homogeneously, and subsequent study tasks and materials were varied according to level of student performance on skills measures.

To summarize, the potential for stigmatizing Title I youngsters exists in both mainstream and pullout settings. However, compensatory education students are less likely to be subjected to labeling and its negative effects in environments where teachers actively encourage children's respect for and appreciation of a variety of human differences; in environments where similarities between learning tasks and materials are emphasized; and in environments where Title I and non-Title I children frequently move, in an organized way, to other parts of the building or classroom to receive special instruction.

Impact of Mainstream and Pullout Instruction on Teachers. Teacher participants at the IDS workshop offered several compelling arguments in favor of, and in opposition to, both varieties of Title I instructional settings. Mainstream settings provided opportunities for the classroom teacher to observe children's academic performance and their reactions to other instructional techniques and adult personalities. In turn, it was believed that observations could lead to increased instructional effectiveness. Workshop participants cautioned, however, that the advantages which observation offers classroom teachers can be confounded by negative factors such as limited or inadequate materials and classroom space, increased recordkeeping, and planning "overload."

Pullout settings offer the advantages of increased communication between classroom and specialist teachers, of providing a model of adult cooperation for children to emulate, and of supporting the child's persevering efforts which lead to increased confidence. Mainstream advocates quickly cautioned that pullout settings are often viewed as intrusions, particularly when the specialist is perceived as a monitor or evaluator of the classroom teacher's practices. Additionally, several workshop participants criticized pullout varieties of compensatory education for interrupting the continuity of instruction and inhibiting "follow through."

Teachers did not express a high level of consensus on the issue. Each of the positions advanced in favor of, and in opposition to, mainstream and pullout compensatory education suggests that the relative effectiveness of either variety is related to teaching and administrative practices.

Characteristics of Well-Coordinated Title I Programs. According to teachers' reports, most well-coordinated Title I programs, whether mainstream or pullout, use similar planning methods and have similar amounts of administrative support for autonomy and flexibility in scheduling.

(1) Planning. Well-coordinated programs seemed to be characterized by joint planning. Much cooperation and communication, both formal and informal, occurred between the regular classroom teacher, specialist teacher, and auxiliary personnel. Additionally, planning that was both systematic and flexible was thought to facilitate matching between classroom instructional events, learner's needs, and district or building goals. This kind of planning also was believed to facilitate consistency between kinds of skills being taught in mainstream or pullout situations and to lead to clear definitions of responsibilities and functions among all adults who contributed to the children's instruction.

(2) Administrative Support for Autonomy. Classroom teachers and specialists who had high degrees of autonomy attributed the success of their respective Title I programs to increased decisionmaking roles. Reported indicators of administrative support included allocation of released time for planning, parity in decisionmaking regarding the employment of auxiliary personnel, and freedom to exercise professional judgment in curriculum matters such as instructional goals and in classroom procedures such as grouping arrangements.

(3) Flexible Scheduling. In well-coordinated programs, Title I children received compensatory academic instruction without being denied participation in other school enrichment activities (music, art, and physical education, for example). Flexible schedules that allowed Title I children to perform successfully in a variety of nonacademic situations reportedly helped to increase student motivation and self-esteem.

Analysis of Practitioner Responses

Theoretical Perspective. The questions posed in the IDS Conference session on setting

have traditionally been the focus of inquiry for organizational and educational sociologists. Considerable attention has been paid to the characteristics of work sites and settings, to repetitive relationships occurring there, and to the delineation and analysis of tasks (Brookover, 1955; Bidwell, 1965; Jackson, 1968; Lortie, 1975; Spady, 1976). The implicit assumption underlying such inquiry is that certain organizational properties of schools and classrooms have implications for the nature, quantity, and quality of teachers' work, teacher-pupil interaction styles, and student outcomes.

Dunkin and Biddle's discussion, "The Classroom as a Social System" (1974), identified several categories of classroom setting variables, including lesson and activity formats, classroom processes and events, and physical characteristics of the classroom environment. After careful review of several substantive reports, Dunkin and Biddle make the following observations regarding effects of classroom features on student achievement:

> Subject matter is found to affect the lesson format, with mathematics featuring a closer, more formal relationship between group function and structure than social studies . . . teacher age and sex also affect format . . . (however) none of the format distinctions . . . has been applied in process-product research. (pp. 209-211)

> Teacher role in classroom processes is also found to affect, or be affected by, a wide variety of other events in the classroom, including lesson format, group function, and classroom location . . . . Altogether, findings for teacher role are more suggestive than those for lesson format, but we have even less evidence of their usefulness in predicting the effectiveness of teaching. (p. 218)

> Roughly a dozen findings report relationships between other kinds of classroom events and pupil roles . . . . Although suggestive, these findings are--once again--as yet unavailable by process-product research. (p. 220)

Regarding other features of the classroom environment, Dunkin and Biddle indicate that "investigators are not in complete agreement on how to conceptualize the physical environment of the classroom" (p. 226).

Critical of the treatment of a diverse array of educational setting variables in previous studies "as separable from one another, linear and additive," Bronfenbrenner (1976) posited an ecological systems view of the educational environment. From his view, the educational environment was conceptualized as a nested arrangement of four successive system structures—micro-, meso-, exo-, and macrosystemic—which account for what happens or fails to happen in an educational setting. For Bronfenbrenner, the educational environment of a learner encompassed both immediate and larger social contexts. Hence, systems properties and processes "affect, and are affected by, the behavior and development of the learner."

When applied to the school environment, Bronfenbrenner's conceptual analysis of the ecological structure of educational environments leads to the following set of generalizations. (1) The learner's immediate setting is the classroom--a place in which the learner engages in particular activities in a particular role for particular periods of time. (2) Interrelations among several sets of settings (classroom, library, resource room, playground, etc.) comprise the learner's instructional setting. At another level, the instructional settings embrace formal and informal social structures that influence, shape, and delimit what learners do within the school. (3) The organizational structures of schools implicity and explicitly mobilize the efforts of teachers, learners, support personnel, parents, etc., in the school environment. These structures, in combination, constitute the school environment and endow meaning to the activity of all persons in the school.

These generalizations are heuristically useful as a conceptual tool for discovering the relevant dimensions of instructional setting perceived by IDS workshop participants. Application of Bronfenbrenner's conceptual model to practitioner-described school environmental features permits a retrospective reconstruction of compensatory education instructional settings.

Classrooms. Group cohesion, and the processes teachers use to develop it, appears to be a classroom characteristic. Teachers use group cohesion to "set the stage" for instruction and learning in compensatory education instructional settings. The concern that teachers expressed regarding breakdowns in classroom peer relationship patterns indicates that teachers perceive a relationship between Title I student achievement and (1) patterns of classroom movement and activity and (2) instructional grouping arrangements. The uniform practices teachers employed in remedying stigmatization suggest that development of group cohesion is an antecedent of effective instruction and learning.

Instructional Process Dimensions. In well-coordinated programs, Title I learners received

instruction which can be described as frequent and repeated over the available time period. The formal and informal cooperation, communication, and joint planning among teachers allowed compensatory education students to study specific skills or concepts in a number of different lesson formats. Joint planning also encouraged considerable repetition. As one teacher explained it, "Our Title I children receive several doses of instruction in a selected content area."

For Title I learners, the instructional setting also appears to be characterized by instructional formats in which skills and concepts are introduced, reinforced, and retaught according to identified performance needs. Ongoing, collaborative monitoring of student progress facilitated the modification and revision of instructional sequencing and pacing.

Well-coordinated Title I programs were characterized by clearly defined areas of responsibility among all adults who were involved directly or indirectly with instructing Title I beneficiaries. IDS teachers perceived clear divisions of job function made according to their responsibility for initiating, reinforcing, and reteaching skills and concepts which facilitated consistency in instructional format. These clear delineations of instructional functions seem to create and maintain content consistency for the Title I learner. By providing uniformity of purpose and direction for Title I instructional efforts, the division of labor provided an accompanying perceptual focus for learner activity and behavior.

School Setting. Key aspects of compensatory education school settings include: (1) released time for planning; (2) administrative support for autonomy; and (3) flexible scheduling.

Released planning time seemed to relate to teachers' opportunities to select and review materials, to assess student progress, and to develop strategies for meeting learners' needs. Equally important was the opportunity provided for collaboration between teachers' aides or Title I personnel and the regular classroom teacher.

Administrative support for autonomy allowed teachers to assume a greater share of responsibility and control in structuring the learning environment relative to curriculum and instruction. Since classroom and subject area teachers performed delineated tasks in the instruction of Title I beneficiaries, it is likely that substantial effort was made to utilize effectively existing temporal, spatial, material, and human support resources.

Teachers reported that schedule flexibility allowed Title I children to perform successfully in a variety of nonacademic classes such as music, art, and physical education, thus contributing positively to the children's increased motivation and self-esteem. Flexible schedules also permitted teachers to utilize instructional time differentially to meet learners' academic needs. Time utilization thus carries implications for the intensity of instruction received by the Title I learner.

Scheduled time prescribes the broad durational parameters in which instruction occurs. Schedules also contribute to the shaping and delimiting of possible instructional arrangements that can occur, such as peer tutoring or lab usage. Flexible schedules provide opportunities for teachers to vary the emphasis placed on certain aspects of instruction at different points along an instructional continuum, e.g., initiating, reinforcing, and reteaching skills and concepts. Several teachers reported that unscheduled timeframes are used to provide additional instruction for learners who experience difficulty. Supplementary instruction allowed learners to complete tasks under supervision and to gain additional practice in specific skill/content areas.

Teachers' comments suggest that schedule flexibility permitted them to vary the amount of time in which a learner is engaged in a given task. Additionally, teachers could vary the quality of supervision provided and exercise greater control over the learning strategies that their students used. Despite absence of time utilization data, workshop participants' reports suggest that intense instruction and considerable instructional continuity exist in compensatory education settings where teachers assume significant responsibility for manipulating temporal features of the instructional setting.

Concluding Thoughts

The preceding setting analysis explored teacher-reported dimensions of compensatory education instructional environments that affect Title I student outcomes. Although the "mainstream" and "pullout" setting dichotomy may have served as a convenient partitioning dimension for reporting Title I student achievement data, the designations by themselves reveal little about the nature of relationships between classroom processes and Title I student achievement. IDS findings may be more useful if viewed as indicators of factors that impinge upon and affect the instructional environment in compensatory education school and classroom settings.

Specific IDS classroom process variables that related to Title I student gains in the 1st grade include: (1) amount of time (reading); (2) curriculum overlap (reading); and (3) assignments and grouping (math). For 3d-grade gains, significant classroom process variables are: (1) amount of time (math); (2) matching by mastery (reading and math); and (3) curriculum overlap (reading and math). Repetitive classroom process dimensions across both grades are curriculum overlap and amount of time.

These reported classroom process findings are compatible with teachers' descriptions of instructional settings. For the practitioner, curriculum overlap related to (1) instructional format and content consistency and (2) amount of repetition that resulted. As noted earlier, joint planning and collaborative monitoring of student progress facilitated teachers' ability to provide a unified instructional program for Title I learners.

Teachers regarded schedule flexibility as a key instructional setting feature. Earlier it was observed that flexible schedules would permit teachers to vary instructional intensity according to learners' needs. IDS findings regarding time allocation are based on interview data concerning regularly scheduled compensatory education instruction. The need practitioners expressed for schedule flexibility strongly suggests that when the parameters of allocated time within a schoolday are sufficiently fluid, other-directed instructional experiences are provided. Varied intensity permits supplementary instruction (i.e., intervention, remediation) to occur when students experience difficulty. Varied intensity can encourage success and increased competence which, in turn, increase student motivation.

It was also noted that flexible scheduling provided opportunities for Title I children to participate in school enrichment activities such as art, music, and physical education. These experiences permitted Title I learners to perform successfully in nonacademic situations and to gain peer acceptance and respect. Teachers noted that the increase in motivation and self-esteem that resulted was positively related to achievement gains.

Assignments and grouping were found to be associated with high achievement only for 1st-grade student gains in math. Workshop participants revealed little that would provide insight into relevant setting dimensions contributing to this phenomenon. Matching by mastery was a uniquely significant factor for 3d-grade student gains in reading and math. Again, workshop participants' responses did not uncover any connections with instructional setting.

The IDS is unique in its attempt to identify and isolate environmental variables that account for student achievement gains in compensatory education. Previous studies failed to uncover process/product connections. The methodological procedures employed in the conduct of the study during the data collection and interpretation/analysis phases has contributed to the discovery of relevant setting features that affect compensatory education student gains.

BIBLIOGRAPHY

Bidwell, Charles E. "The Social Psychology of Teaching." In N. L. Gage (ed.), Handbook of Research on Teaching. American Educational Research Association. Chicago: Rand McNally, 1963.

Bronfenbrenner, Urie. "The Experimental Ecology of Education." Paper presented as the AERA Award address at the Annual Meeting of the American Education Research Association, San Francisco, 1976.

Brookover, Wilbur. A Sociology of Education. New York: American Book Company, 1955.

Dunkin, Michael J., and Bruce J. Biddle. The Study of Teaching. New York: Holt, Rinehart and Winston, 1974.

Jackson, Phillip. Life in Classrooms. New York: Holt, Rinehart and Winston, 1968.

Lortie, Dan C. School Teacher. Chicago: University of Chicago Press, 1975.

Poynor, Hugh et al. Instructional Dimensions Study: Final Report. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1977.

Spady, William G. "Power, Legitimacy and the Process of Evaluation." Unpublished monograph, May 1976.

TOWARD FUTURE STUDY OF THE EFFECT OF SCHOOL SETTING ON LEARNING

Roger W. Shuy
Georgetown University and
The Center for Applied
Linguistics
Washington, D.C.

The past decade has witnessed an increasing sensitivity to the broad range of competencies which are developed by children as they learn in the school setting. Setting is a complex concept, involving the verbal and nonverbal behavior of children and others which accompanies the perceptual, physical, and social properties of the people, objects, and events involved in that behavior. For example, the teacher, the child, and the aide are all interacting in various role relationships in a classroom during a given schoolday. These factors all contribute to the learning environment of the child. The importance of setting in educational matters might be seen, in fact, as analogous to the importance of context in language study. For years, linguistic work has focused on the universals and the broad generalities which characterize a language. Only within the past decade, however, have linguists begun to examine the context or the setting in order to understand the variability of language to a fuller extent (Hymes, 1974). Linguists realized that this variability brought richness to language expression and alternative ways of understanding and producing nuances, and that however important the generalities and universals are, our humanity is clearly revealed through our variability. It is the use of this variability, in fact, which allows for individual expression, apt phrasing, and even poetry. The development of sociolinguistics in the past decade or so has been evidence of the growing concern for setting in educational issues. It is not surprising, then, that education is concerned with the effect of setting on learning. We have always known that context contributes heavily to children's development. It is only recently, however, that we have begun to specify the dimensions of context enough to begin to see how it actually works.

It is with this background that the IDS exercised a first-cut view of the effect of school setting on learning. As Bloom has pointed out, we cannot decide between alternative interpretations of the same behavior without a systematic analysis of context. Gumperz and Herasimchuk, in their study of teacher/student interactions, show that children make use of a number of variables such as task expectancy, role differences, and previous utterances in the conversation to formulate interpretations and, thereby, to learn. It is not difficult to extend this list of variables related to setting to include factors related to classroom placement, size, competing or simultaneous activities, accoutrements, equipment, and so forth.

Although setting has been viewed in many ways, depending on the objectives of the observer, a systematic way of differentiating the possible perspectives called setting has not been made. Therefore, efforts to bring about policy-oriented research which utilize setting as a focused dimension may prove difficult. Some relationships between setting and children's learning will be direct. Others will be indirect. Some will affect some children but not others and not always in the same ways. Such variables are extremely difficult to control, and even when such control is attempted, one is left with the observer's paradox--that by controlling for the variables, the researcher distorts the reality of the experience to the extent that the research can be invalidated. The study of setting is motivated by its interpretative, explanatory, and predictive power. Unless we understand children better and guide their education more effectively as a result of such study, we cannot consider such work as helpful.

It was out of understandings such as these that NIE undertook to determine the effect of one major aspect of setting on the learning of certain children. Setting is only one part of a larger category of influences on learning which we shall refer to as context. Four different types of context have been recognized in the literature: physical, situational, social, and linguistic. These categories are by no means mutually exclusive, and most of the recent work on language context, for example, sees it along with (rather than exclusive from) physical and situational contexts (Scollen, 1976). Context is not a fixed set of properties of the world which children take into account as they learn. Cook-Gumperz and

Corsaro (1976) consider context as part of the communication and learning process, a set of fluctuating variables which are constantly being reevaluated by all participants in the learning process during the interaction.

Any policy implications growing out of the study of the effect of school setting in learning need to consider three very important questions:

(1) Is there agreement on what it is that is being measured?

(2) Does the research measure the right things?

(3) Does the research assess at a definable and appropriate point of measurement potential?

This paper discusses these three issues with the hope that by being clear on these aspects of educational research, we stand a better chance of making progress toward assessing the setting of learning in such a way that policy decisions can be made effectively. It also suggests direction for future research. It is agreed that setting is an important variable which contributes to or detracts from learning. If effective policy decisions are to grow out of such study, we need to discover how setting both helps and hinders learning. It is the thesis of this paper that academic researchers have, to date, made little progress in assessing the effect of setting of education because they have neither described it fully nor examined it deeply. Consequently policy decisions will be based, necessarily, on a rather shaky research base.

Is There Agreement on What Is Being Measured?

One reason why policy implications have been difficult to make regarding the effect of setting grows out of the very nature of this activity called research and evaluation. Such activity is, of itself, unnatural, and it imposes compromises of one type or another. We have already mentioned the problem of the observer's paradox. Even purely observational study upsets the ecology of the classroom, instituting new and often difficult constraints. Alternatively, research and evaluation can carry out isolating types of studies in the following way:

(1) Stop life long enough to look at it and determine what is going on

(2) Decide what to look at (also a decision about what not to look at) and what to consider as similar and what to consider as different

(3) Determine how much is enough to matter

These two procedures, observational and isolating, constitute evaluation circles. The former has the advantage of naturalness of data; the latter has the advantage of ease in analysis. Both must understand the dimensions of the topic well enough to know what to look for and do it at an appropriate time.

Unfortunately, policy-oriented research seems not to have the luxury of waiting for all the details which academic researchers feel are necessary. To fully understand the context of learning (of which setting is one component), one should know a great deal about the dimensions of that context either to observe them in natural contexts or to isolate them for study exclusive of such contexts. Put even more simply, this merely says that in order to measure something, one first has to know what it is that is critical to measure. Since policy decisions must be based on such evidence, such decisions will be no better than the evidence upon which they are based.

In an effort to discover the effect of context on learning, the IDS research examined one aspect of the physical property of context. This aspect was identified as pullout vs mainstream instruction; that is, supplemental instruction delivered to students outside of the regular classroom vs supplemental instruction delivered inside the regular classroom. This definition seemed harmless enough until the issue was considered outside of the physical dimension alone. What do people perceive this contrast to be? What social events supersede or overlap its measurement? These questions would not be asked in a study which isolates for measurement one property by itself.

Any comparison of mainstreaming with pullout classes is difficult to understand when there is no broadly agreed-upon definition of these terms. If the terms were placed on a continuum, at least general agreement might exist concerning the polarities of meaning, but the middle aspects will be vague and unspecified. Regardless of the definitions used for the purposes of the research, the 40 teachers assembled to discuss this project had their own concepts of what pullout means. To some of them, any child taken from the regular class constituted a pullout. At the other extreme, one teacher reported that the whole class, including the teacher, was pulled out to the resource teacher's room.

Despite the general definition used in the IDS, it should be clear that in the field there is not unanimous agreement as to which is which. Perhaps even more startling is that the teachers interviewed often praised or blamed one or the

other of these setting strategies on bases not germane to the issue. Labels play an extremely important role in education, and it appears that if we tell some teachers that they are in one of two different modes, many are willing to believe it, whether or not the modes are in any way accurate.

One thing obvious in the Instructional Dimensions Study is that the researchers and the teachers seemed to have a great potential for seeing the settings in quite different lights. It is likely, in fact, that researchers would also not necessarily agree with each other about what it is they are measuring. A great danger in evaluation is in assessing a feature without defining it precisely and without seeing it in an ecological balance. The IDS is no more guilty in this regard than is the broad field of knowledge. The state of the art has not yet been adequately developed to assess this feature competently.

Does the Research Measure the Right Things?

Assuming that researchers can agree on the features being measured, an even greater question is the "so what" issue. How do we know that the now agreed-upon and well-defined feature really matters? Of all the things one might assess, why choose this one?

Assessing the effectiveness of any national program involves great risk taking. The question of what to assess often gets short shrift in educational evaluation and may fall prey to the more visible and easily measurable (countable) issues. The setting of education seems to be of tremendous importance, but it is certainly much larger and different in scope from the mainstream vs pullout dimensions noted here. Any science seeks explanations for events and behavior. These explanations are often found in the variables which are involved. The variables involved in education include the topic of instruction, the learners, the setting of instruction, the teachers, and the delivery system itself (techniques, materials, curriculum, etc.).

Even though the setting of instruction is a critical variable, it cannot be seen in isolation from its co- and subvariables. In order to determine exactly what these other variables are within the variable of context, one must accomplish a number of tasks. Pullout vs mainstream, as a category, masks many variables which can confound whatever results one might obtain by comparing these two features in isolation. Even some classroom observation instruments, as weak as these are, include several variables related to setting (attractiveness of the room, size, the nature of the seating, height of ceiling, lighting, etc.). These features, in themselves, are enough to cause binary comparisons between pullout and mainstream to be less than useful (see for example, Jane Stallings).

In the case of reading, the setting is confounded by a number of other factors. For one thing, reading instruction, as Griffin (1977) points out, is by no means limited to that portion of the day labeled "reading." By this she does not mean only that "every teacher is a reading teacher" or that there is reading involved in the content areas. Instead, she indicates from her research that comprehension is taught in such contexts as when teachers read stories to their students when the children do a cooking experiment. To measure reading ability, then, and to ascribe gain or loss based on whether or not the child was mainstreamed or pulled out overlooks a great deal of learning which is irrelevant to either variable.

Likewise, recent research by McDermott indicates that a great deal of learning which is not easily identified as reading is accomplished in reading groups in elementary schools. Children learn about turn taking (actually a very complex phenomenon with very subtle signals), how to know whether or not their efforts have been approved, how to seek clarification when they do not fully understand, and many other important social skills (McDermott, 1977b).

An effort to evaluate reading instruction in a given classroom, therefore, will need to include more than the reader's achievement. It will need to observe and record the conditions which help or hinder the reader's development. Recent ethnographic studies of the classroom have revealed some very interesting things. (For a comparison of quantitative and qualitative assessment procedures, see the entire issue of Anthropology and Education Quarterly, May 1977.) Such observational procedures stand in stark contrast to most evaluation efforts involving national profiles in education (McDermott, 1977a).

A somewhat different perspective to the importance of the classroom setting is offered by Cicourel (1974). Noting the importance of the literature on the limitations of attention, memory, and types of recording procedures in most research, Cicourel observes:

> If we can believe the importance of short-term memory in human information processing, then the consequences of a classroom lesson for learning by disadvantaged children are especially serious. If short-term memory is a viable notion then new information being received requires some sort of rehearsal, and the capacity of the memory becomes central because of the amount and complexity of incoming information and its possible

> displacement of other elements. A . . . tester would have to be familiar with conditions (1) in which certain types of acoustical and visual memory deteriorate; (2) in which conditions can influence the subject's ability to process information provided initially by standardized American English instructions that are acoustically and partially visual; and (3) in which the instructions require a translation of verbal material into visual operations or standardized verbal operations. (p. 328)

Cicourel goes on to observe that current social and behavioral science research places too much emphasis on the analysis of verbal materials detached from the setting and occasion of use and relies too much on the restricted occasions such as formal tests. The displacement of visible and imagined objects and events can be reconstructed by some people much more efficiently than others. Unless these more subtle and difficult-to-study issues of social interaction are examined carefully, researchers can easily fall back on a social pathology or genetic explanation of differential school performance (Cicourel, 1977, p. 332). If the children try to make use of what they considered to be common knowledge of the setting, and this knowledge is not the common knowledge of the school, they stand to be judged unfairly.

To agree to measure a component of setting, then, research must first identify a wide range of subvariables and then determine which ones support and which ones confound the major variable and finally assess the extent to which this effect can be taken. An example of this sort of error can be seen in the recent national evaluation of bilingual education via Title VII programs (American Institutes for Research, 1978). This evaluation took a very global view in which the actual theoretical differences between programs were never identified. As a result, programs were compared on the basis of unknown and potentially irrelevant features. The results were useless to the goal of providing guidance to Congress for future directions. It was assumed that the setting was being measured, but no effort was made to define the possible dimensions of setting, and no subvariables or theoretical positions were considered.

The point here is that any evaluation of instruction is based on several possible theoretical stances. A decision to measure early learning skills in isolation from each other or from the gestalt of learning is a strong, theoretical decision which is subject to critical evaluation. The results of such measurement are only as good as the theory which undergirds it. Likewise, a holistic measurement will tell us little about the early developmental skills which preceded it. One should not be confused with the other. Both should be evaluated for how they can help us understand the acquisition of learning.

A consistent and appropriate learning theory is not the only necessity for a useful evaluation of the effect of setting. Also critical is a consistent and appropriate theory of the subject being taught. In the case of reading, for example, it is necessary to know the content of reading, not just the methodology of delivering it to children. The content of reading is composed largely of linguistics, psychology, and anthropology. Reading is a language-processing operation (linguistics), learned (psychology) within an overlapping set of cultural constraints and influences (anthropology). To evaluate such a phenomenon as reading effectively is to know more than the traditional knowledge (or misinformation) about reading. Much of what passes for traditional knowledge in this field is close to folklore. Great gaps of knowledge are glossed over as though we really know the answers.

In addition, any focus on the setting of education needs to consider the different sets of behavior and discourse rules which are in operation. These sets stem from all of the "nonsetting" variables such as age, sex, socioeconomic status, ethnicity, and geographic region, as well as the more relevant educational setting conditions such as room size, place, and other more local environmental conditions. The entry of evaluation into a normal classroom setting creates the observer's paradox: if we do not observe, we cannot tell what is going on; if we do observe, we upset the ecology of the classroom behavior.

Are We Assessing a Definable and Appropriate Point of Measurement?

To make an assessment of the influence of setting on learning, it is necessary to do it at an appropriate stage of development. To measure reading ability in the first month of schooling, for example, would be ludicrously ill-timed. Likewise, to assess the early developmental walking skills of a mature normal adult would seem foolish. There are clearly inappropriate points for evaluation.

In order to avoid inappropriate points of assessment, it is necessary to have either clear and undisputable research evidence as a touchstone or, barring such evidence, it is necessary to have a reasonable and well-defined developmental theory.

For example, since reading is a language-processing operation, a theory of reading should primarily account for language in the many ways it reveals itself developmentally. In at least one

sense of the term, reading can be viewed in a timeframe continuum in which early skills are developed, paradoxically, only to be abandoned as soon as possible for advanced cognitive processes. Reading offers a rather clear example of a mixture of such early behavioral skills, later cognitive strategies, and a potential for cultural interpretation and individual learning style. Thus the setting of measurement can be seen to change radically from point to point on a timeframe continuum. Early reading skills, such as letter-sound correspondences and word-part decoding, are clearly worth measuring at the time when such learning is going on. But they are early developmental skills which, when a reader later becomes proficient, become so automatic that the reader is not even conscious of them. It would seem ludicrous to measure such automated skills at a later point of development. In fact, one might even predict that such skills, no longer at the level of consciousness, might appear to decline on tests which try to measure them. Good readers do not read letter by letter; they proceed by larger and larger units, up to and including discourse meaning units.

Learning theory has long held that different kinds of learning can take place at different stages of learning. Thus behavioral, skill-focused learning can be adequate in the early stages of reading but should be replaced as soon as possible by more cognitive strategies which involve higher level meanings. For some critics, this means sentences. I would argue that for the typical cooperative child who will accept the fact that certain dull or odd things must be done in order to get to more interesting things, almost any reading approach can be successful. Some children may be ready to accept such behavioral (letter-sound type) instruction earlier or later than others. Some sort of diagnostic instrument should be able to predict who such children are. It seems likely that one type of prediction will be based on the child's personality more than on the reading tasks or on the language accesses themselves.

Parallels to this exist in other areas of learning as well. Foreign language learners face similar problems. At a given stage in the development of the target language, the learner experiences different degrees of awareness or automaticity in different relationships to each other. Thus in his communication, a learner at onset will be expected to maintain high awareness of phonology and lexicon which is automated and therefore less useful to measure at later stages.

However much the experts might argue about the exact levels of such automaticity, there is general agreement that something like such a relationship actually exists, and this is the major point to be made from the illustration. Any effort at assessing learning ability must be fully aware of this aspect of the setting or run the risk of measuring useless things.

Suggestions for Future Research

These three basic questions, then, underlie the Instructional Dimensions Study of the effect of setting on learning. The goal of this aspect of the study was rather precise, focusing as it did on one small aspect of the larger notion of context. The following diagram is only suggestive of the dimensions in which the effect of context could be assessed:

Figure 1 simply pictures the potential gross categories or features which might be noted as children learn. Duplicate figures could be presented for learning and for teaching, doubling the dimensions and realistically noting that teaching and learning are not the same things. The IDS examined one cell in the column which notes physical properties. It did not attempt to see the social or perceptual properties, to distinguish behavior dimensions or, perhaps most critical of all, to determine the interactive, dynamic effect of one dimension on another. Being policy research, the IDS had to build on what is known. Figure 1 clearly demonstrates how little this amounts to.

		PROPERTIES		
		Perceptual	Physical	Social
Behavior	None			
	Verbal			
	Non-Verbal			

FIGURE 1
THE CONTEXT OF LEARNING

Policy research, depending as it does on academic research, must demand a better basis from the academic world if its conclusions are to be optional. This paper is not a criticism of policy research per se, but of the conditions which detract from its effectiveness. Perhaps one of the best outgrowths of policy research is the weakness it discovers in that knowledge base. Perhaps one of the best services it can perform is to point out these weaknesses.

The IDS and the follow-up conference for 40 of the Title I teachers revealed a number of topics for which we need to know a great deal more.

The Effect of Autonomy on Setting. Teacher judgments about the setting of compensatory education, specifically mainstreaming and pullout approaches, are interesting if not explanatory. On

the whole, the 40 teachers studied favored the pullout setting. The reasons used to support pullout instruction were varied in quality and relevance.

One pro-pullout teacher explained that teacher autonomy was the reason she preferred this approach: "Pullout works very well--I wouldn't have it any other way. Mainstreaming couldn't work--my kids don't even know when I leave. In my program we do our own budgets, sit in on hiring interviews, and state our preferences. Autonomy is the key."

Another teacher felt that when the pullout teacher takes away her children, she loses control: "I don't know exactly what he is doing with ending sounds."

One approach to studying the context of learning would be to study the notion of teacher autonomy. If the preference for pullout classes stems from the teacher's desire to be in control and to not have other teacher equals (or superiors) in the classroom, the meaning of the concept of pullout is considerably broadened. This is essentially a study of teacher attitudes, security, and locus of control. Do teachers with high quotients of autonomy do better jobs in one physical setting than another? This is a clear case of support for the idea that the properties of context not only are broader than the physical dimension but also are interactive and interdependent. In short, it is necessary to study the process in a dynamic rather than a static framework.

The Effect of Physical Properties on Context. Teachers need materials to work with. Using the analogy that if a little salt is good, a lot is better, some Title I teachers seem to believe that their children need more nonbook materials. One of the conference teachers said: "So many Title I children need manipulatives. It would be extremely difficult to do this in a mainstream environment." It is not at all clear why manipulatives or equipment such as tape recorders cannot be used in a mainstream setting, but this response met general agreement.

On the other hand, some teachers objected to the overprecision and lock-step of materials geared to Title I children. One objection to the pullout regimen had to do with the lack of creativity which is perceived to be involved with the pullout instruction. Said one teacher, "I like to have a more creative atmosphere."

What is the appropriate relationship of teaching materials to the Title I context? Is the salt analogy true? Do compensatory education materials develop a lock-step dullness? This aspect of the physical/social/perceptual context of learning has certainly not been clarified. Every teacher's convention is flooded with manipulatives, plastic gadgets, and other nonbook substitutes. Do we really have a clear idea of their value? Is market research adequate? Can it replace learning theory or avoid it? Since most manipulatives are related to early skill learning, is it not possible to compare high-manipulative and low-manipulative physical contexts? What effects do such contexts have on the perceptual and social properties of context? Can the work on the stimulus characteristics of younger children be extended to school-aged learning? (Nelson, 1973; Clark, 1978).

The Effect of Pullout on the Subject Missed During the Pullout Experience. The 40 teachers surveyed in the IDS observed that pullout students missed physical education, science, social studies, learning centers, art, and music. Oddly enough, few teachers, except for mainstream supporters, seemed to lament this loss. It is difficult to know what this situation portends for the total learning of a child, but it is certainly measurable. Here the perceptual and physical properties of context come together in a most interesting way, and it would be of great benefit to education to learn what this interaction yields. Careful documentation of curriculum in a longitudinal study would be desirable.

The Effect of Peer Teaching and Stigmatization. The supporters of mainstreaming argue that peer teaching does go on and that in mainstreamed classes peer stigmatization is greatly reduced: "Some of my pullout children become very belligerent. Some kids say they are dumb."

In response to a question concerning how peer learning could take place when those in need were pulled out for special instruction, another teacher responded, "There's enough of the rest of the day to do the modeling." One teacher actually conceptualized the role modeling issue as one of contrasting theories. She put it this way: "One approach is that kids learn best if they have the same ability. The other is that kids of different abilities learn best from each other." This was a profound, though simple, observation. It is absolutely necessary for research to address this issue. We need to know a great deal more about the effect of the social properties on conceptual and perceptual development. No assessment of the context of learning will be complete without taking this issue into consideration.

Styles of Learning. Many teachers had opinions on (or near) the topic of grouping. One expressed her feelings as follows: "The more adults you have in the room, the smaller the group, the better the teaching." Her theory is

clearly that small is good and that children learn better in small groups than in large. There is little or no evidence to support or reject this notion. Relatively little is known about learning style, despite the rather large amount of attention given it in education. We know (or think we know) some things about cultural learning styles (Navajos sit in circles, etc.) but relatively little about individual, nonsocially determined styles. Some research exists on perceptual styles, and experience shows us that in judging a work of art, some people see shape before color or detail. It would seem reasonable that such perceptual plugging-in might be relevant here as well. It would seem reasonable to expect research to learn how to assess the effect of learning in groups versus learning by oneself both cross-culturally and individually so that teaching programs might adjust to such useful knowledge.

The Effect of Coordinative Ability. It is clear that some teachers prefer to be the lone teacher in the classroom. Others thrive on team efforts. Little is done in teacher training (or in academic research, for that matter) to develop the notion of team effort. The term is used, but it is little more than a slogan.

Pullout teachers who had experienced the mainstream setting complain about the lack of coordination between the regular teacher and the compensatory education teacher. One teacher asked, "How can an outsider know my kids?" It remains unclear how this same outsider can know her kids any better when the child is pulled out rather than mainstreamed. The real problem appears to be the old "two cooks in the kitchen" syndrome. When asked whether or not an aide in the regular classroom offers a similar threat, one teacher responded: "It's hard for two people to work in the same room. You've got to be able to respect each other. You can't always have your own way." Other teachers disagreed, saying that a teacher and aide in the same room did not present a "two cooks" problem. Again, authority appears to be an important issue. If pullout programs are implemented primarily to avoid the "two cooks" issue, education had better take a good look at itself. Surely teachers must learn to work together if they plan to help children learn to work together. More importantly, teacher training institutions might be expected to do a better job of training teachers in coordinative practice, whether with equals or subordinates.

The Effect of Recordkeeping. When pullout instruction is introduced, complications in record-keeping seem to develop.

One of the teachers in the IDS was particularly opposed to the increased amounts of recordkeeping involved in such efforts: "I don't think it's worth it--my teacher judgment is just as valuable as a card with x's on it." Another teacher noted that the time spent keeping records kept her from interacting with the students: "I feel that I don't meet the kids' needs because I have to get the paperwork for skills work done." This paperwork not only involves keeping track of where the children are, but it is also attached to accompanying sets of "skills-hierarchy materials" associated with many pullout programs. Some teachers lamented that their children were taking an inordinate amount of time getting through these hierarchies and some of them seemed to be able to read anyway. When asked why they needed to teach and test the skills even after the learning gestalt had been accomplished, the teachers offered as their only reason that it was useful to have documentation in record form when justifying grades and placement to parents, principals, and even to the children themselves. This rather legally oriented fear needs to be further examined. It appears at least possible that pullout instruction, requiring more bookkeeping, can actually detract from learning even though it may protect the teacher or school. Other ways of protecting might be considered if learning suffers because of it. Once again, the total context of learning must be examined if we are to understand it in its fullest sense.

Comparative Setting Studies. In order to fully understand the effect of setting on learning, one might expect to be able to measure the same learning in different settings. One of the odd egocentrisms of education is that learning is directly related to what is taught. Some specialists in language teaching have begun to realize that much more language learning goes on outside of classrooms than inside. The social properties noted in Figure 1 are the socially recognized events both within and outside the school setting. The setting has been shown to affect the way children interpret the utterances of others, either literally or nonliterally. "I see someone whose hands are not folded" can be taken literally, for example, but is unlikely that the setting in which this is uttered will permit a mature, literal understanding. Setting also affects the way children behave and talk (Bates, 1976). Education is only beginning to understand this dimension, and it would be useful to know a great deal more. Policy research on learning will be somewhat handicapped by the lack of basic research in this field.

Conclusion

The task of evaluating the setting of learning, then, is of considerable complexity. Setting is more than plus-or-minus place. Far more critical than this is to identify what it is that one wishes

to measure with such precision so that some agreement will ensue about what is being measured. Next, one tries to determine whether or not this variable really matters. Then one searches for the appropriate time and place to measure it. Much of what we examine when we measure reading ability has no more bearing on a child's ability to read than does a driver's ability to remember the fact that he is supposed to look in the rear view mirror every 10 seconds bear on his ability to drive a car. In driver's training, one is taught to look into the rear view mirror every 10 seconds, but to be consciously aware of doing so could easily lead to a traffic disaster. An early learning skill, however useful, is not necessarily a good measure of overall ability or competence.

At this time, policy research will do well to demand that academic research provide a better basis for determining the effect of setting on learning. We need to define what it is that we mean by setting, set its measurement within appropriate content theory models, place it within appropriate and well-defined learning theory models, and know its qualitative aspects before we attempt any kind of quantitative studies. The teachers surveyed in the IDS seemed to have a feel for these problems even though they were never articulated in quite this way. We need to see learning in its natural setting. Even policy research needs to identify the theory upon which it is based in order to avoid mixing theories and calling them the same thing. We need to identify a measurement variable in relationship to the effect of many other contributing variables. The teachers themselves suggested such obvious variables as classroom size and attractiveness and teacher concern. One could suggest differences in culture, age, sex, ethnicity and socioeconomic status.

We might begin by following the lead offered by researchers in microanalysis (see, for example, Kendon et al., 1975; Erickson, 1975; Sheflen, 1973). In order to arrive at a description of how children make sense of the world, it is necessary to examine in detail what these children are doing and saying. Just as people "style switch," depending on what they perceive to be required of a situation (Blom and Gumperz, 1972), so students and teachers act differently in the classroom, depending on the situation in which they are engaged. If such variability takes place (and it does), it must signal something very important about learning. In any case, a great deal of careful observational and ethnographic work will need to precede the counting which characterizes much educational evaluation today. To assess the effect of setting on learning, we will need to expand the dimensions of the topic to determine what we mean by setting, what really matters in terms of measuring, and where or at what point on the learning continuum it is best to observe it. Meanwhile, policy decisions will have to be made on less than adequate grounds.

BIBLIOGRAPHY

Bates, E. "The Development of Conversational Skill in 2, 3, and 4 Year Olds." Pragmatics Microfiche 1(2), 1975; Language and Context. New York: Academic Press, 1976.

Blom, J. P., and John Gumperz. "Social Meaning in Linguistic Structures." In J. Gumperz and D. Hymes (eds.), Directions in Sociolinguistics. New York: Holt, 1972.

Bloom, L. Language Development: Form and Function in Emerging Grammars. Cambridge, Massachusetts: MIT Press.

Cicourel, Aaron V. "Some Basic Theoretical Issues in the Assessment of the Child's Performance in Testing and Classroom Setting." In A. V. Cicourel (ed.), Language Use and School Performance. New York: Academic Press, 1974.

Clark, E. V. "Building a Vocabulary." In P. Fletcher and M. Garman (eds.), Studies in Language Acquisition. Cambridge, England: Cambridge University Press, 1978.

Cook-Gumperz, J., and W. A. Corsaro. Social-Ecological Constraints on Children's Communicative Strategies. Berkeley Language Behavior Research Lab Working Paper No. 46, 1976.

Erickson, Frederick. "One Function of Proxemic Shifts in Face-to-Face Interaction." In A. Kendon et al. (eds.), Organization of Behavior in Face-to-Face Interaction. The Hague: Mouton, 1975.

Griffin, Peg. "How and When Does Reading Occur in the Classroom?" Theory into Practice 16(5):376-383, 1977.

Gumperz, J., and E. Herasimchuk. "The Conversational Analysis of Social Meaning: A Study of Classroom Interaction." In M. Sanches and B. Blount (eds.), Sociocultural Dimensions in Language Use. New York: Academic Press, pp. 81-115.

Hymes, Dell. Foundations in Social Linguistics. Philadelphia: University of Pennsylvania, 1974.

Kendon, Adam, et al. (eds.). Organization of Behavior in Face-to-Face Interaction. The Hague: Mouton, 1975.

McDermott, Ray P. "The Ethnography of Speaking and Reading." In R. W. Shuy (ed.), Linguistic Theory: What Can It Say About Reading? Newark, Delaware: IRA, 1977.

McDermott, Ray P. "Kids Make Sense." Doctoral dissertation, Stanford University, Stanford, California, 1977.

Nelson, K. "Some Evidence for the Cognitive Primacy of Categorization and Its Functional Basis." Merrill-Palmer Quarterly 19:21-39, 1973.

Scollen, R. Conversations with a One Year Old. Honolulu: University of Hawaii, 1976.

Sheflen, Albert. Communication Structure: Analysis of a Psychotherapy Transaction. Bloomington, Indiana: Indiana University Press, 1973.

Chapter V. Planning, Organization, and Management

ASPECTS OF PLANNING, ORGANIZATION, AND MANAGEMENT OF SELECTED READING PROGRAMS

Morton Botel
University of Pennsylvania
Philadelphia, Pennsylvania

Local Schools Make a Difference

The focus of this report of the Instructional Dimensions Study (IDS) Conference is on aspects of planning, organization, and management of the reading program from the perspectives of the 40 conferees and from related research literature.

Conferees were selected teachers in compensatory education programs of the IDS. They shared their experiences and insights on the planning, organization, and management of their school district, school, and classroom instructional programs in reading.

The research on teacher effectiveness (Heath and Nielson, 1974) has not documented a single stylistic teacher skill which can clearly be associated with student achievement in reading. Nor have major methodological studies shown the value of one reading method over another as having great significance (Dykstra, 1968; House et al., 1977). There is, however, considerable empirical evidence that certain variables in planning, management, and organization can enhance teacher effectiveness in advancing reading achievement. These variables are a balanced reading/language arts program, a simple management system that involves more holistic, functional testing, strong principal leadership, and parent involvement.

The thesis to be explored in this report is that local schools do make a difference in producing higher achievement in reading, but the difference seems to be more a result of motivational variables and of planning, organization, and management of the total reading/communications arts environment than of stylistic variables of teachers or of the reading "method" used.

Statements of Conferees

The 6 to 7 hours of dialogue with the 40 conferees provided insights into the contributions and blocks to learning of the leadership staff of their schools, into their own practices and preferences, and into the growing role of parents.

The conferees were most critical of reading goals set forth from the central administration. They believed the goals tend to be too broad and abstract. Further, they thought the administration tends to view accountability from the top down, expressing expected performance of students in terms like "to improve each student's achievement in reading each year by at least one year on the xyz test."

Typically, they said, the means for implementing the main goals are specified by tests, management by objectives schemes and by texts and other resources that have highly specified skill sequences. These are chosen by the top of the school administration hierarchy with little teacher input, although there is a small but increasing amount of teacher input in the case of choosing textual materials. It was the consensus of these teachers that the districtwide management and testing schemes result in a heavy burden of nonproductive overtesting and paperwork.

They also noted that such experiences as reading to children, self-selected reading, and oral and written composition are typically not found in the statements of main goals or in specification of means for achieving them.

The conferees considered the principal's leadership very important. While only about half of the conferees considered their principals to be strong, they all agreed that principals who took active interest and leadership in planning, monitoring, and conferring with teachers elevated the quality of reading instruction.

The teachers indicated by their descriptions of their work that they were responsive and flexible, aware of the individual differences and needs of children. They felt free to modify and often to ignore management and test data in favor of their own observation of children in learning settings. They reported spending 3 to 4 hours per day to prepare for their classes. Despite the lack of

district-level articulation of such critical learning experiences as reading to children, self-selected reading and composing, the teachers typically built them into their programs. Moreover, they tended to coordinate their work with that of the regular classroom teachers and their aides.

The teachers noted that parents are increasingly involved in advisory and aide functions and in helping their children at home in reading and arithmetic. They reported that their schools are sending appropriate materials home and instructing parents in their use.

Inferences from Conferee Statements for Improvement of Reading Programs

Four main inferences can be drawn from the above observations of the conferees for improving the planning, organization, and management of school reading programs.

First, the district management scheme should not be limited to abstract goals and/or to stating expected scores on tests. It should articulate the insights and practices of teachers who provide children with a balance of productive holistic reading/language arts experiences involving oral literature, self-selected reading, and oral and written composition as well as systematic skill sequences.

Second, the district management scheme for monitoring mastery of basic decoding/comprehension skills should be simplified, making use of more holistic or functional measures of reading competence, thus cutting down the seemingly endless and unproductive paperwork and record-keeping of typical management plans. It should assert the significance of responsive/diagnostic teaching as the main evaluative approach to meeting children's learning needs. This is in fact what good teachers do, but it is given little credibility in district plans.

Third, the principal should be a strong manager of the school reading program by articulating the structure and management of the program as indicated above and by providing for dialogue with and among the staff and parents.

Finally, parents should be encouraged to become partners with teachers in the direct education of their children in the basic skills, and schools should help them find the means for doing so.

Support for These Inferences from Research

This section of the report will examine the research support for the inferences derived from conferee observations. It will suggest that a balanced program of reading/language arts involves providing students with four critical experiences, rather than a particular best method, that competency testing should be more holistic rather than atomistic, that principal leadership can have a strong influence on reading instruction, and that a literate environment at home is a powerful variable in developing a competent reader.

A More Functional (Holistic) Approach to Teaching: A Balancing of Reading/Language Arts Experiences

The conferees agreed that it would be helpful if the district goals and means delineated and supported a balance of all of the critical reading/language arts experiences, rather than concentrating only on those that related to specific subskill sequences in decoding and comprehension. In addition to working with structured sequences or skills, the conferees generally provided regular times for children to hear stories and poems, to have them respond in ways that stimulated oral and written language and imagination. They also provided time for their children to choose books and read them silently. They regularly enabled children to dictate and write narratives and poems. It is important to remember that they did so in the absence of districtwide policy encouraging these practices. Many colleagues of the conferees may be influenced by the lack of stated district policy to think that experiences other than basals and other hierarchically organized programs are purely for enrichment, and are therefore not "basic" skills.

The evidence indicates otherwise. Each of the above experiences is critical to success in reading because each contributes to achievement in reading as well as to the formation of positive attitudes, according to a review of the research by Botel (1977), who defined them as follows:

Critical Experience 1: Responding to Literature. Students experience literature by hearing it, reading it, or viewing it and are encouraged to respond to it through dialogue, writing, simulations, and the other expressive arts.

Critical Experience 2: Sustained Silent Reading of Self-Selected Books. Students choose books, periodicals, etc., from a wide selection and read them without interruption at their own pace and in their own way.

Critical Experience 3: Composing, Oral and Written. Students dictate or write something of their own each day.

Critical Experience 4: Investigating and Mastering Language Patterns. Students study system-

atic sequences of decoding/comprehension activities at their instructional levels through both problem solving and practice and drill approaches.

A More Functional (Holistic) Approach to Testing and Recordkeeping

Half or more of the conferees found the testing and recordkeeping in their schools to be cumbrous and excessive. Without exception, however, they reported that they relied mainly on their own observation and judgment to provide for the special needs of each child. An increasing number of schools have recently adopted testing, management, and recordkeeping systems that focus almost exclusively on subskills. In such schools, reading is analyzed into subskills of sound, structure, and meaning, each of these isolated fragments having a test and learning routine associated with it. Some management schemes identify hundreds of these subskills.

The negative effects of such systems are many. The schemes define a reading curriculum as the sum of these pieces; they demand an inordinate amount of time for testing and recording children's performance; they distract teachers from the balance children need of the critical reading/language arts experiences.

In the past few months, the National Council of Teachers of English (NCTE) and International Reading Association (IRA) have spoken out strongly against such excessive focus on testing and subskills. From the perspective of these two professional organizations, the effects have been to diminish the student's comprehensive ability to read and write, to narrow and to dehumanize the curriculum, and to dictate teacher style. These groups urge teachers to teach from a variety of perspectives and to use a variety of learning routines. What they disapprove of is trying to measure formally, using either standardized or criterion-referenced tests, the innumerable perspectives of meaning as though they had strong diagnostic value. In short, there seems to be a different conceptual framework for instruction and for formal testing.

Now, to the issue of schoolwide testing on standardized and criterion-referenced measures. Which of the many perspectives suggested by this taxonomy can be monitored by district- or schoolwide tests?

Before dealing with this issue, it ought to be made clear that the responsive teacher, who is aware of the perspectives of meaning and structure and who knows routines for engaging students in thinking about them, is also constantly evaluating informally how well students are dealing with these ideas. In this sense teaching and evaluating are two sides of the same coin.

Should standardized and criterion-referenced tests try to evaluate each of the perspectives of meaning and structure suggested by the dimensions and aspects of a comprehensive curriculum such as the one appended? The answer must be no for several reasons: cost effectiveness, fragmentation of the curriculum, and unreliability. Administering tests that are extensive enough to deal separately with all the important aspects of meaning and structure would be prohibitively expensive and time consuming. On the other hand, if we simply choose the aspects we like or select them randomly, we run the usual risk that the curriculum itself will follow from the selection, and a comprehensive program will have to shrink.

In terms of fragmentation, the nature of tests is such that it suggests to some that a curriculum is the sum of its elements. In fact, learning is a function of understanding relationships and interactions, of analyzing and reorganizing content, not of compiling and mastering a series of discrete elements. Therefore, even if it were possible and cost effective to test everything that a curriculum embraces, we might be misled into thinking that performance on separate items represented high-level thinking skills rather than memory.

In terms of reliability, all the evidence on comprehension subskills indicates that our present measurement of them with standardized or criterion-referenced tests is not reliable enough for individual or group diagnosis. There are two general conclusions from all correlational studies (there are no experimental studies of this question) of the distinctiveness and hierarchical nature of measurable comprehension subskills of reading.

First, there are between one and four significant factors in comprehension, according to several reanalyses of Davis's (1969) data (Thorndike, 1973; Spearitt, 1972). According to Thorndike, 80-90% of the variance in the Davis studies of components of comprehension is accounted for by vocabulary and the remainder by two or possibly three factors. According to Spearitt, ". . . present types of reading comprehension tests, as distinct from word knowledge tests, largely measure one basic ability corresponding to the label of reasoning in reading" (Spearitt, 1972, p. 110).

Second, there is no statistical evidence supporting a hierarchical order for sequencing comprehension instruction (Davis, 1972, p. 172).

The logical and comprehensive (taxonomic) analysis of perspectives of reading, then, yields

many dozens of perspectives. The statistical analysis renders one major factor and between one and three others. The inferences of these facts for the role of formal tests of comprehension are clear. The richness and diversity of comprehension cannot be described by subtests. Tests can provide a global estimate of general comprehension, but responsive imaginative teaching is the only productive way to know how students are thinking.

Following the suggestions made above for districtwide testing would be cost effective, reducing the time and money for testing and recordkeeping to less than 5% of what many school systems are presently using. At the same time the results would be more reliable and therefore more informative. Educators could then emphasize the continuous process of developing a comprehensive instructional program with its many instructional perspectives and its countless opportunities for responsive/diagnostic teaching.

The Principal as Leader

The conferees, whether they regarded their principals as effective leaders or not, agreed that a strong principal could be most influential in helping develop a strong reading program.

Two U.S. Office of Education studies yielded support of the conferees' observations (Dykstra, 1968; Abt Associates, Inc., 1977). These studies analyzed the effectiveness of various combinations of published primary reading curricula. In both studies the major finding was that local schools make a difference--that labels of method or model make small contributions to achievement compared with the uniqueness of individual schools, neighborhoods, and families.

In commenting on the 1968 report, Carroll and Chall (1975) emphasize the principal's importance.

> One of the more interesting and illuminating findings in the U.S. Office of Education (USOE) Cooperative Reading Studies was that some schools and school systems had consistently better results than others, even when the types of communities were similar, and when similar methods and materials were used. The chief factor that seemed to be responsible for the superior results in these schools was the amount of interest and attention given to the organization of the reading program by the school administrator. Teachers of reading need understanding and encouragement from their superiors--supervisors, principals, school superintendents, and school board members. They need to work in a climate that supports and rewards their efforts, gives them help when they falter, and allows them freedom to try out new approaches that seem reasonable to them. (p. 23)

Parent Involvement

Conferees spoke of the growing involvement of parents in helping their children learn how to read. They reported a recent increase in the number of workshops for parents and in the number of helpful materials that are routinely sent home.

Considerable research has shown that one of the main variables in reading achievement is the "literate environment" at home. Children who come from an environment in which they are read to and in which they are encouraged to read on their own become more school-wise, book-wise, and test-wise (Thorndike, 1973). Several studies have verified the significance of parent involvement in reading at the preschool and later stages (Levenstein, 1975; Sakamoto, 1975; Sprigle, 1972; Chomsky, 1971, 1972; Durkin, 1974).

Carroll and Chall (1975) strongly urge that parents play a stronger role in helping their children learn to read:

> . . . we feel that parents should be encouraged to give their children help in reading before (and after) they enter school--by reading to them, by teaching them such things as the names of the letters of the alphabet and the sounds of the letters, and by helping them learn to recognize a small vocabulary of words. Parents should give it only if a child seems to show an interest in letters and words, and in reading generally--interest that parents can promote by showing that reading is a pleasurable and meaningful activity. We are aware that, on the basis of widely held but poorly supported views in the reading profession, reading teachers have frequently discouraged parents from giving help in reading on the grounds that it would "confuse" the child and produce conflict with teaching in school. We believe this is unlikely to happen. (p. 19)

Summary

In short, the Instructional Dimensions Study conferees' perceptions of what is productive for children's reading competence corresponds to inferences from research.

In the conferees' view, it was important in their programs to have regular periods of reading to pupils, to provide time for self-selected reading and self-selected writings as well as for the study

of specific decoding and comprehension skills. They believed that there was too much testing and unproductive recordkeeping in most districtwide management systems, and they relied mainly on their own judgment in meeting the needs of pupils. They thought that strong supportive principals made an important difference in the morale and quality of the program. They generally urged parent involvement in helping their children learn to read.

With respect to these same variables, the research literature suggests that:

(1) Children would benefit from a systematic yet balanced program of four critical experiences in reading/language arts. They are:

   a. daily opportunity to hear and to respond to a variety of imaginative prose and poetry selections;

   b. daily opportunities to select books and read silently at their own pace;

   c. daily opportunities to dictate and write their own thoughts and feelings; and

   d. daily opportunities to investigate and master the functional skills of decoding in the search for meaning.

(2) The management and monitoring program needs to rely on more simple, reliable, and cost-effective functional reading tests, and on informed teacher judgment. Such management would minimize the excessive focus on subskills and its concomitant time-consuming recordkeeping, and would provide more time for direct instruction.

(3) Strong principal leadership is associated with high achievement in reading. Major studies to date do not support one method over another as a major variable in achievement. On the other hand, differences among schools that use the same method are profoundly significant. The well-organized, knowledgeable, supportive, and enthusiastic principal seems to make the difference.

(4) Informed parent involvement results in higher reading achievement. No variable is more consistently associated with achievement than the quality of the literate environment at home. Experiments have validated this regularly observed relationship. Where parents have been helped to establish a more literate environment, their children's reading achievement has been enhanced.

Clearly, teachers perceptions correspond to the findings of research in the matter of improving the reading achievement of children. What is missing very often is the understanding of and commitment to these ideas by the leadership of our schools.

BIBLIOGRAPHY

Abt Associates, Inc. Education at Experimentation: A Planned Variation Model. Cambridge, Massachusetts, 1977.

Bailey, G. M. "The Use of a Library Resource Program for the Improvement of Language Abilities of Disadvantaged First Grade Pupils of an Urban Community." Unpublished doctoral dissertation, Boston University, 1969.

Berg, P. C. "Evaluating Reading Abilities." In W. H. MacGinitie (ed.), Assessment Problems in Reading. Newark, Delaware: International Reading Association, 1973.

Bettelheim, B. The Uses of Enchantment. New York: Alfred A. Knopf, 1976.

Bissett, D. J. "The Amount and Effect of Recreational Reading in Selected Fifth Grade Classes." Unpublished doctoral dissertation, Syracuse University, 1969.

Bogart, M. Paperbound Books in New Jersey Schools. Trenton, New Jersey: State Department of Education, 1965.

Bormuth, J. R. "Defining and Assessing Literacy." Reading Research Quarterly 1:7-66, 1973-74.

Botel, M. Pennsylvania Comprehensive Reading/Communication Arts Plan. Harrisburg, Pennsylvania: The Pennsylvania Right-to-Read Plan, Department of Education, 1977.

Botel, M., and N. Botel. A Critical Analysis of the Taxonomy of Educational Objectives. Washington, D.C.: Curriculum Development Associates, 1975.

Bowen, B. E. "The Teacher's Role in Teaching Literature in the Elementary School." Unpublished doctoral dissertation, University of Michigan, 1964.

Broening, A. M. Developing Appreciation Through Literature. Studies in Education, No. 13. Baltimore, Maryland: Johns Hopkins University Press, 1929.

Carroll, J. B., and J. S. Chall (eds.). Toward a Literate Society. The Report of the Committee on Reading of the National Academy of Education. New York: McGraw-Hill Book Company, 1975.

Carver, R. P. "Reading as Reasoning: Implications for Measurement." In W. H. MacGinitie (ed.), Assessment Problems in Reading. Newark, Delaware: International Reading Association, 1977, pp. 44-45.

Chall, J. Learning to Read: The Great Debate. New York: McGraw-Hill Book Company, 1967.

Chomsky, C. "Write First, Read Later." Childhood Education, March 1971.

Chomsky, C. "Stages in Language Development and Reading Exposure." Harvard Educational Review, February 1972, pp. 1-33.

Clay, M. M. "Early Childhood and Cultural Diversity in New Zealand." The Reading Teacher, January 1976, pp. 333-342.

Cohen, D. "The Effect of a Special Program in Literature on the Vocabulary and Reading Achievement of Second Grade Children in Special Service Schools." Unpublished doctoral dissertation, New York University, 1966.

Cohen D. "The Effect of Literature on Vocabulary and Reading Achievement." Elementary English, February 1968.

Davis, F. B. "Research in Comprehension in Reading." Reading Research Quarterly 4:499-545, 1969.

Davis, F. B. "Psychometric Research on Comprehension in Reading." Reading Research Quarterly 7:628-678, 1972.

Durkin, D. "A Six Year Study of Children Who Learned to Read in School at the Age of Four." Reading Research Quarterly, fall 1974.

Dykstra, R. "Summary of the 2nd Grade Phase of the Cooperative Research Program in Primary Reading Instruction." Reading Research Quarterly, fall 1968.

Educational Experimentation: A Planned Variation Model, Vol. IV--A-D. Cambridge, Massachusetts: Abt Associates, Inc., 1977.

Elkind, D. "We Can Teach Reading Better." Today's Education, November-December 1975, pp. 34-38.

Fader, D. N., and E. McNeil. Hooked on Books: Program & Proof. New York: Berkley Medallion Books, 1968.

Farr, R. Reading: What Can We Measure? Newark, Delaware: International Reading Association, 1969.

Gibson, E. J., and H. Levin. The Psychology of Reading. Cambridge, Massachusetts: MIT Press, 1975.

Graves, D. Balance the Basics: Let Them Write. Durham, New Hampshire: Education Department, University of New Hampshire.

Guth, P. "A Study of the Characteristics of the Fourth, Fifth, and Sixth Grade Pupils of Two School Districts Who Are Underachieving in Reading." Unpublished doctoral dissertation, University of Pennsylvania, 1976.

Haught, E. H. "Students' Patterns of Thinking in Teacher-Led Group Discussions and Student-Led Small Group Discussions of Literature." Unpublished doctoral dissertation, University of Illinois, 1973.

Heath, R. W., and M. A. Nielson. "The Research Basis for Performance-Based Teacher Education." Review of Educational Research, Vol. 44, No. 4, 1974.

Hession, C. "An Examination of Methods of Having Beginning Readers Acquire Basic Sight Vocabulary Through the Use of Recommended Picture-Story Books Incorporating a Multimedia Approach." Unpublished doctoral dissertation, Columbia University, 1973.

House, E. R., G. V. Glass, L. D. McLean, D. F. Walker, and E. J. Hutchins. No Simple Answer: Critique of the "Follow Through" Evaluation. Urbana, Illinois: Center for Instructional Research and Curriculum Evaluation, University of Illinois, 1977.

Hunt, L. C. "Can We Measure Specific Factors Associated with Reading Comprehension?" Journal of Educational Research 51:161-172, 1957.

Jackson, C. C. "An Evaluation of the Effect of a Special Reading Program Adapted From Aesop's Fables on Reading Achievement and Self-Concept of Fifth Grade Students." Unpublished doctoral dissertation, Boston University, 1974.

Levenstein, P. The Mother-Child Home Program. Freeport, New York: The Verbal Interaction Project, 1975.

MacGinitie, W. H. "What Are We Testing?" In W. H. MacGinitie (ed.), Assessment Problems in Reading. Newark, Delaware: International Reading Association, 1977, pp. 35-43.

Michalak, J. "Head Start-Type Programs Get Second Look." New York Times, April 30, 1978, p. 9.

O'Hare, F. Sentence Combining. Urbana, Illinois: National Council Teachers of English, 1973.

Pfau, D. W. "An Investigation of the Effects of Planned Recreational Reading Programs in First and Second Grade." Unpublished doctoral dissertation, State University of New York at Buffalo, 1966.

Porter, E. J. "The Effect of a Program of Reading Aloud to Middle Grade Children in the Inner City." Unpublished doctoral dissertation, The Ohio State University, 1969.

Raftery, Sr. F. "The Effect of a Quality Literature Program Conducted by Elementary Education Majors on the Reading Achievement of Second Grade Students." Unpublished doctoral dissertation, New York University, 1974.

Rosenshine, B. Skill Hierarchies in Reading Comprehension. Monograph, Center for the Study of Reading, University of Illinois.

Sakamoto, T. "Preschool Reading in Japan." The Reading Teacher, December 1975.

Schneeberg, H. "The Listen-Read Program: A Four Year Study." Unpublished study, School District of Philadelphia.

Serwer, B. L. "Linguistic Support for a Method of Teaching Beginning Reading to Black Children." Reading Research Quarterly, summer 1969.

Sirota, S. "The Effect of a Planned Literature Program of Daily Oral Reading by the

Teacher on the Voluntary Reading of Fifth Grade Children." Unpublished doctoral dissertation, New York University, 1971.

Spearritt, D. "Identification of Subskills of Reading Comprehension by Maximum Likelihood Factor Analysis." Reading Research Quarterly 8:92-111, 1972.

Sprigle, H. The Learning to Learn Program. ERIC (ED/066/669), 1972.

Strickland, D. S. In Carl Senna (ed.), Can Slum Children Learn? The Fallacy of I.Q. New York: The Third Press, 1943.

Strickland, D. S. "The Effects of a Special Literature Program on the Oral Language Expansion of Linguistically Different Negro Kindergarten Children." Unpublished doctoral dissertation, New York University, 1971.

Thorndike, R. L. Reading Comprehension Education in Fifteen Countries. International Studies in Evaluation III. New York: Halsted Press, 1973.

Thorndike, R. L. "Dilemmas in Diagnostics." In W. H. MacGinitie (ed.), Assessment Problems in Reading. Newark, Delaware: International Reading Association, 1977, pp. 57-67.

Wanat, S. F. "Evaluating Reading: An Introduction." In S. F. Wanat (ed.), Issues in Evaluating Reading. Arlington, Va.: Center for Applied Linguistics, pp. 5-11.

Weber, G. Inner City Children Can Learn to Read. Washington, D.C.: Council for Basic Education, 1972.

Wilmot, M. P. "Toward Reading of Elementary School Students Including in the Reading Program a Period of Sustained Silent Reading." Unpublished doctoral dissertation, University of Colorado, 1975.

REPORT AND COMMENTS ON NIE TEACHER WORKSHOPS, INSTRUCTIONAL DIMENSIONS STUDY

David Hawkins
University of Colorado
Boulder, Colorado

I would like in this paper to turn attention to the substantive aspects of the NIE Instructional Dimensions Study--in the very incomplete form so far known to me--and to my personal records and impressions of the panels I participated in, entitled "Planning, Organization, and Management of Compensatory Education," during the April 27-28 conference.

General Context

Ours is a field in which the main problem is not to seek more information, but rather to know how to redesign the channels of information flow--by successive approximation--so as increasingly to admit what is more essential for practical or theoretical purposes and to filter out what is inessential.

My concern in this paper is not the study itself--which I hardly as yet know--nor the teachers' conference itself--from which I have a crowd of somewhat contradictory and in any case very interesting impressions--but the development of some possible and plausible relations between the two which might both benefit the interpretation of data from the former and make possible some reconstruction of the channels through which other relevant information might be obtained.

The general impression I gathered from the teachers I observed is that the dimensions of teacher concern with the study were in general only loosely related to those evolved for purposes of research and sometimes were almost orthogonal to the latter. This applies particularly to the teachers' discussion of planning. We learned from all that the assumption of stability was largely justified; there were indeed program changes over the past 2 years in some cases, but the changes were not of a kind to perturb the teachers' sense of confidence or to be beyond the ingenious remedies they in some cases invented. The essential stability was that of the teachers themselves, on the whole an experienced group. This impression was sustained in all the later group discussions.

Some New Variables

In the first of the four groups, we experienced a change of style in the latter part of the discussion, one that suggested the order that we followed in later groups. The pivotal questions had to do with planning. There were preliminary comments that made evident a diversity of planning histories with respect to the kinds of programs and the kind and degree of district and teacher involvement; what brought the meeting to a distinctly higher level of liveliness was the opportunity to focus on teachers' own perceptions of their essential role, as teachers, in planning for the children they taught.

The teachers spoke enthusiastically of their real job--making children comfortable; making them want to come to school; taking time from a prescribed lesson to discuss a recent hurricane in the neighborhood and find out more about such things; replacing a narrow "diagnostic-prescriptive" impasse with specific children by materials improvised by the teacher herself in the light of a quite personal diagnosis. A general summary--confirmed when it was mentioned in later groups--was that where teachers were committed to following the pattern of some specifically adopted curricular "package," they exercised substantial freedom and judgment in interpolating materials from other sources, including those they made themselves or asked parent groups' help in manufacturing.

The same sense of independent mind and judgment was apparent when we touched on the matter of assignment of children to the Title I category. Teachers started with the test, but they felt fully confident in their ability to modify the indication of test scores in the light of personal judgment about individual children--whether or to what extent they carried this through being only a matter of the degree of flexibility accorded them. They got more relevant information from observing children take the test, in some cases, than from the score itself. Some children are conspicuously test-shy, others are test-wise without the abilities their scores

imply; entering 1st graders have often not learned the game.

I should not like these observations to be interpreted as evidence that our teachers belonged to a class of rebels, as advocates or practitioners of radically nonstandard teaching modes such as "open education," etc. With two or three exceptions, they did not seem to be. They were mostly stayers who accepted their present-day school world as given; within it, in one way or another, they functioned with considerable competence.

A significant turnaround in the later discussions came with a question asked by my co-reporter, Morton Botel. This concerned teachers' actual and preferred priorities with respect to their allocation of children's time between various possible uses of the total allotted to reading instruction. His report discusses the specific responses we received. Generally speaking, the teachers in the three groups were strongly in favor of giving top priority to free reading and related activities calculated to involve children in reading as a personal interest and commitment, along with decoding practice. On the other hand, some confessed they did not follow this belief in practice, primarily because of time pressures--in effect, the routines of decoding instruction had a higher priority. In the discussion, Morton Botel asked whether this kind of activity was in their district plans, and typically it was not.

Unfortunately, time pressures were such that I was able in only one group to ask the same sort of question about time spent in classroom math or "math labs." I received a similar reply. From observation elsewhere, I would conjecture that top official priority is typically given to work with narrow skills in the written, workbook mode, and that a more investigative style of work, for instance with "three-dimensional materials"--geoboards, cuisenaire rods, poker chips, geometrical tiles, cubes and multibase blocks, and balances, to name only the most stylized--is relatively unknown or unused. Even less likely is the use of practical arithmetic and geometry in measuring, mapping, model building, or recordkeeping. I would conjecture that any teachers' tendencies toward exploration and use of such materials may be inhibited by unfamiliarity or by the absence in district or school plans of any recognition of their value, either for children as a whole or for those in the Title I category.

I should like to propose here a rather general hypothesis about teaching and planning in general, and Title I in particular. I shall not limit the hypothesis to planning for Title I children; I regard most elementary school teaching of reading and math as essentially in the same category.

Hypothesis: In stating this hypothesis, I shall use an analogy that I owe to Frances Hawkins, an experienced therapeutic teacher of young children, taken from her book The Logic of Action (New York: Pantheon-Random House, 1974, pp. viii-x). The analogy: Oxygen given to save the lives of premature infants was for long being given in what turned out to be excessive amounts. The excess caused blindness and other brain damage. Such results, however, are characteristic products of lack of oxygen, of anoxia, and this fact seemed at first to indicate a need for more oxygen rather than less. After several years of hospital research it was finally discovered that a high oxygen pressure damaged the delicate lung tissues and thus, after a time, decreased their absorptive capacity for oxygen, causing retrolental fibroplasia, damage to the late-stage embryology of the brain, notably that of its optic nerve. I quote Frances Hawkins' use of the analogy:

> Faced with the failure of children in our schools, their failure to learn well along the track which school has paved for them, where are the school doctors (not from the outside) who will say, with such informed and persistent conviction, "It is something we are doing to them, our schools are doing to them?" Instead of seeing a child's failure as a response to our doing, to our failure, it becomes a "learning disability," a "behavior problem," and we are exonerated.
>
> Very much of what children need for their learning must come directly and indirectly from adults. As oxygen to the lungs, it must be readied for them and transmitted to them. Faced with failure in the process we respond too easily by increasing the intensity of the efforts which have already failed, and in doing so we may block the very channels through which children gain knowledge and understanding. . . .The input we offer is needed, yet not assimilated.

Title I children are predictably (for whatever reason) not likely to assimilate well the standard fare of early schooling. Under these conditions increased drill, increased emphasis on decoding and competency skills, as distinguished from the contextual aspects which make these skills desirable and desired by children themselves, can be, like excessive oxygen to the lungs, a therapeutic agent transformed into a poison, the unwitting cause of that which it is aimed to cure.

I refer to this general statement as a hypothesis, which from the NIE research perspective it must be--although I believe its correctness is a guiding commitment among the most successful of our teachers. It runs counter to many present-day

tendencies toward commercial "packaging" of detailed curricula and implies instead that teachers are or should be able to adapt and to create an ambience in which children's curiosities are aroused and supported. This implies no lack of attention to the mechanics of "skill," but rather a context in which children are motivated to practice these skills and to learn them along with things of greater intrinsic value.

In both reading and mathematics these considerations of background and ambience are of vital importance. Since my own work has been primarily in the latter domain, let me indicate in more detail what I see as being involved. Even before starting school, almost all children have a wide range of mathematical competencies and understanding related to their preschool and out-of-school lives; in contexts that invite their close attention, they can communicate verbally their awareness and grasp of a wide range of basic ideas of number, form, and order. But this body of fluent understanding is not where we typically start in school "mathematics." School mathematics is focused on the mastery of a code, itself an incomplete and abbreviated written language, often bewilderingly detached from the context in which children acquire fluency of early mathematical understanding. Reliance on prepackaged "curricula," even more than the less detailed textbook approach of the past, necessarily ignores all elements of context which establish continuity with earlier or concurrent informal learning. The little-step-by-step approach of the workbook, the diagnostic test, and the prescriptive loop back through exercises constitute an unwitting conspiracy to cause the very failure they purport to cure.

I say all of this with full appreciation of the ultimate importance of the step-by-step character of formalized mathematics itself. Even in its most elementary branches, mathematics is knitted together into an intelligible system by components of a detailed logic which is the final touchstone. But no one grows in fluency and mastery by paying attention only to those components. The more long-range elements--those of practical experience, imagery, of analogy, of intuition--are equally vital to children's learning.

The above hypothesis is not in "testable" form. It is a conditioning hypothesis, a directive framework for more detailed formulation. I have no adequate reason to believe that the participants in the conference would go all the way in agreeing with my formulation. I would surmise that a good many of them have not had the opportunity to sample across a range of learning situations much wider than the one that schools normally provide--which is not very great. Nevertheless, I believe they would go part way. They would move beyond the dominant views expressed in state and district policy and planning. Their stability as professionals who do experience the constraints I have objected to creates a research opportunity.

Possible Research Opportunities

The opportunity for research exists in the rather wide gap between standard demographic research and that based on attempts to create learning environments that are demographically too rare to be available for formal investigation. Attempts to bridge this gap by experimental designs of the kind recently popular are significantly unimpressive. I need not refer to recent controversy over the comparison of "models." The essential point is that the time scale required for the appearance of any kind of program worth studying is long compared to what most research designs allow. Interesting school situations are the product of institutional and personal history, not of short-term engineering.

"Stability," one of the IDS criteria, can help create interesting school situations. There has been time allowed for teachers to settle in, to invent ways of working with children, to bring their background expertise to bear, to carve out for themselves some domain in which they have relative freedom in their planning and day-to-day decisionmaking.

Suppose one chose the top 10% and the bottom 10% of classes, measured by average gains in the study's test scores. One would then look for contrasting characteristics of the two groups of programs thus selected, particularly with respect to more detailed definition of those characteristics mentioned above--notably the characteristics of teachers' participation in planning at various levels and on various time scales, the constraints they experience, and their perception of the relative importance of various aspects of their program.

As I have already stated, I would be surprised if schools and teachers selected by such a 10% criterion would differ radically from the majority in most surface aspects. Even with due regard for regression and for the narrow sampling of educationally significant learning represented by standardized tests, I would, however, expect that some valuable results could come from such a second round.

Most large-scale educational research that I am familiar with starts with the attempt to measure and relate predefined and inexpensively sampled variables that are suspected or claimed to have relevance, and the typical outcome is one of relatively minor differences in average outcomes. It seems to me that such investigations

can be significant only as leading to a second or third cycle in which the primary value accorded to earlier results is the redefinition of what is looked for in the following cycle.

Let me illustrate this possible redefinition by an example of one of the matters discussed in the conference and highlighted in its preliminary report, the lack of conspicuous difference between "mainstream" and "pullout" patterns of association between classroom teachers and specialist teachers. There appears to be a wide variety within each of these patterns, and this has a great deal to do with the degree and content of communication between the teachers involved. It is possible to have a "pullout" pattern in which classroom and specialist teachers supplement each others' perceptions of the needs of individual children, having seen them in different environments. On the other hand, it is possible to envision a "mainstream" pattern in which there is little or no such communication, where the visiting specialist is in effect a total "pullout."

Rather plausibly, there is a definition of alternative patterns in such cases that should take a bigger bite of the variance in outcome measures than the one initially adopted. The information is likely to be more expensive per unit to obtain, which argues for a reduction in the scale of research--but the information does not seem to be inherently difficult to quantify through a combined use of interviews, questionnaires, and directed observations.

What I am proposing is a shift from evaluative research to elucidative research. The one looks for association between input and output variables already defined and makes comparisons between two or more alternative patterns of schooling. The other starts with significant differences in educational outcomes and tries to define essential differences in treatment. This means that the empirical investigation is relatively open, guided by hunches, by analogies, by extrapolation of theory, as well as by known and obvious requirements.

But the contrast between evaluative and elucidative should be for interaction, not for separation. To define new variables of input and outcome that are more strongly related to each other and can be understood within a conceptual organization is also to learn to define better channels for evaluating by this new evidence, for further elucidation, etc.

Having suggested one example, I would like to make a series of suggestions. Since I have by no means fully grasped the framework or content of the study, I must apologize for what will undoubtedly be seen as a lack of contact with the kinds of information it already contains.

Let me first make a specific suggestion concerning future outcome measures. In any subsamples selected for further empirical study, one should somehow obtain "mini-longitudinal" data on children's later performance in school--test data, teachers' judgments on individual children's general welfare and promise in school, absenteeism, etc. Lacking a 20-year longitudinal study of educational outcomes, we could very well use a second year, and a third, especially when test measures can be qualified by teacher judgments, etc. There may well be other input-outcome measures to be used, indirectly but perhaps partially related--such as, for example, evidence concerning continuing stability, teachers' morale, etc. All the new measures would seem to me to be relevant as follows:

1. Opportunity. As I understand this composite measure, it is lacking in any qualitative dimensions. "Overlap" relies exclusively on tests as outcome measures. If other outcome measures are added, the educationally spurious value of "overlap" would presumably be reduced, as it would on tests explicitly designed to avoid thoughtless literal repetition of content taught.

2. Individualization. This term has come to be used, in the curriculum-package industry, for what is generally its opposite, namely the so-called diagnostic/prescriptive. Clearly, in the IDS individualization measures, an effort has been made to avoid this sloganized meaning of the term; the attempts to gather data about assignments and grouping, alternative learning routes, and sequencing would, I hope, provide a basis for further scrutiny of opportunity actually afforded children for learning in ways adapted to individual talents.

The teachers' own observational skill and inventiveness with respect to supporting children's already achieved strengths is, I believe, the best clue I can suggest for analytical guidance. This, however, is not easy information to come by in any crisp form, and the role of theory can only be to suggest what kinds of information to seek. IDS researchers' attempts to discover the degree to which teachers create unique assignments for individual students and the extent to which teachers try to improve the sequence through use of supplemental materials are steps toward obtaining this information.

In the whole long tradition of "remedial teaching," the anoxia analogy gains in plausibility from the fact that prescriptions and provisionings are almost universally thought of as based on the diagnosis of weaknesses, not on the use of children's attained strengths and coupling with the motivations which undergird those strengths. For example: a 2d-grade teacher with seven children

held back for a year from a previous teacher decided to take them entirely off of the standard reading fare for most of the term. Instead she substituted for them a science corner to which they contributed, with much ado about labels; treasure hunt games with written instructions; favorite story books read to them and then read by them, etc. Two weeks before test time she "put them to the book" again, for test-taking practice. They all passed. This is only a surface description of a teacher's art in the weaving together of reading with other interests found in these specific children, but it makes the essential point.

3. Instructional Events. I do not have sufficient information to comment intelligently about specific variables under these headings or to know where to focus. The overwhelming majority of them seem to leave out one obvious dimension: conversation among children and adults who share in work significant to them jointly. Informal conversation can of course be idle--but in the presence of worthy and absorbing materials and task, it is probably the chief instrument of educational rapport.

4. Motivational Processes. In the context of the sort of variables I would seek to define, this category is not as a whole distinguishable from the previous one. One could find ways of assessing the classroom climate in terms of use of children's versus teachers' wall displays, the presence of three-dimensional materials available for use, the degree of simultaneous diversity of activities (and passivities) and day-to-day maintenance or change of pace, and the quality of classroom sound and motion. The most important new variable might be a teacher's capacity to orchestrate ground rules, freedoms, and choices made available.

5. Teacher Background. In addition to the kinds of information already categorized, it would be important to know from teachers themselves what factors or episodes or histories in their own background they consider to be most relevant to their current philosophy and practice. In particular, it would seem important to direct their attention, in this matter of background, (a) to the presence or absence of serious work in (normal) child development, and to their judgment about the value of such study; and (b) to the nature of their own subject matter education.

Summary

The nature of the IDS and the use of teacher consultants in it may provide a basis for further analysis of existing data and for added empirical research.

From the point of view of the school careers of children and of elucidative research, the crucial variables are very partially reflected in the study's major input and outcome variables. "Overlap" is of some importance, but it needs disentangling as to whether it means something like instructionalized coaching "for the test" or wider significant learning that a test can reliably sample.

The responses of teacher conferees strongly suggest that there be a specific further focus on matters directly affecting children's response in school--on classroom ambience, on teachers' abilities to see needs in relation to learning strengths, and on their repertoires for meeting these needs when so seen.

I have suggested that one should look at reasonably small samples at the extremes of what are, admittedly, educationally unreliable outcome measures in order to find patterns--complexions of recorded or still-needed variables--that seem to be more essential than those one has, perforce, started with. The aim of this research should be explicitly exploratory--not to prove anything, but to suggest desirable redefinitions of information-gathering categories.

The plasticity of children's learning potentials is far greater than the evidence from demographic or I.Q. data usually indicates. The reason is obviously that situation sampling--sampling of the kind that emphasizes the importance of context and situation in giving children access to their own resources for learning--is as difficult as the really interesting situations are rare.

The moral is that one should, for scientific value, throw out large-group averages for the time being and concentrate on the relatively rare extreme cases. Reading and math tests are geared to prevailing averages and distributions, and when read as meaning more, expose us to what should be called the demographic fallacy. When teachers really tap children's curiosities and help them become seriously involved in learning over a substantial period, they can show 2- or 3-year gains in the narrow skills. According to the demographic fallacy, such gains are several standard deviations from the average, and thus to be discounted. But in our field, pay dirt is also by definition rare.

Such small-sample exploratory research can only, at first, provide existence theorems--not causal relations, and certainly not uniqueness theorems, if there are any. But by closer observation we may learn to develop some useful theory--better operational definition, more carefully constructed information channels--and thus hope finally, from the study of larger samples, to create public credibility and understanding for major directions of improvement.

Chapter VI. Individualization of Instruction

IS IT TRUE WHAT THEY SAY ABOUT INDIVIDUALIZED INSTRUCTION?

Marianne Amarel
Educational Testing Service
Princeton, New Jersey

Introduction

In the spring of 1975 the National Institute of Education (NIE) embarked on a comprehensive evaluation of compensatory education programs supported by Title I of the Elementary and Secondary Education Act of 1965 (ESEA) and state funding. Compensatory programs encompass a wide variety of services intended to reduce or overcome the differences in educational attainment commonly found among children with differing socioeconomic backgrounds. The evaluation was undertaken in response to a Congressional mandate directing NIE to conduct a study of purposes and effectiveness of compensatory education programs so as to provide information relevant to future legislative decisions regarding primary and secondary education.

The stipulation for evaluation has been built into federally funded compensatory programs since their inception in the mid-1960's. The 1974 Amendment to ESEA authorizing NIE to examine the programs serving compensatory education students reiterated Congressional expectations that evaluations be used to inform policy decisions. Mandated areas of study ranged from an examination of ways used to identify children needing compensatory programs to an assessment of how effectively the children's needs have been met. These directives defined the boundaries of the study in general terms only; they spoke to the means and strategies of implementing it not at all. The challenge of formulating the study questions about the critical and discriminating dimensions of a multipurpose and widely spread program was quite properly left for NIE to meet.

A Short History of Large-Scale Evaluations

During the past several years, numerous evaluations of educational programs, varied in scope and aims, were undertaken. The results of large-scale evaluations proved to be disheartening on at least two counts: they failed to demonstrate measurable, consistent, or notable effects on student attainments that could be linked to educational, particularly compensatory, programs; they were equally unsuccessful in relating specific features of educational contexts to measures of effectiveness. The first study to attract national attention, the Westinghouse/Ohio evaluation of Head Start, now appears methodologically naive, yet the application of increasingly sophisticated tools has not yielded more definitive or less controversial results (White et al., 1973; SRI, 1971; Wargo et al., 1972; Dyer, 1978).

Attributing differences in the outcomes of schooling to the influence of measurable dimensions of school settings has proved to be equally elusive. In an analytic review of studies of the impact of school resources on students, Spady (1976) finds only equivocal results. Drawing a distinction between tangible resources and "value climates," he directs attention to an aspect of schools that has resisted measurement and has thus been absent from most impact studies, much to their detriment. Evaluations in search of program-related effects have not fared any better. The national evaluation of Follow Through Planned Variation (Stebbins et al., 1977), which did report some sponsor-related outcomes, is currently under challenge (House et al., 1978). Scores of other studies could be cited that failed to identify dimensions of schooling consistently associated with students' educational attainment.

The history of Title I evaluations, thoroughly documented by McLaughlin (1975), adds up to a chronicle of unproductive and, in retrospect, often misguided activity. When the assumptions of the input/output model that guided these evaluations are closely examined, the model emerges with damaged credibility. The suitability of an essentially technological model for evaluating diverse decentralized programs has recently been questioned on multiple grounds (Berryman and Glennan, 1978; House, 1978). But, in fact, its feasibility has been doubtful for some time, given our limited capacity to conceptualize and measure

educational input and output. Instructional programs, labeled but largely unexamined, have been typically designated as the input, or as independent variables, and standardized test scores have served almost exclusively as the indicators of effectiveness, or output. Evaluations unable to detect program differences have thus been open to criticism for the crudity of their measures, while studies that did report differential effects lacked the finer descriptors needed for relating differences to program characteristics.

Not only the critics but also the designers of impact studies have pointed out methodological, conceptual, and political problems frustrating their efforts. Campbell (1975) listed a few under the general heading of "meta-scientific issues," indicating the variety of problems constraining program evaluation:

> . . . there is a precarious rigidity in the measurement system, limiting recorded outcomes to those dimensions anticipated in advance; process is often neglected in an experimental program focussed on the overall effect of a complex treatment, and thus knowing such effects has only equivocal implications for program replication of improvement; broad-gauge programs are often hopelessly ambiguous as to goals and relevant indicators; changes of treatment program during the course of an ameliorative experiment, while practically essential, make input-output experimental comparisons uninterpretable; social programs are often implemented in ways that are poor from an experimental design point of view; even under well controlled situations, experimentation is a profoundly tedious and equivocal process . . . (pp. 8, 9)

On the whole, then, the yield of summary evaluations has been thin. In the absence of a suitable theoretical and methodological base for identifying and measuring critical dimensions of schooling and their effects with any confidence, the studies did not provide information useful for policy decisions. Their utility for educational practice was equally meager, so little relation did they bear to the reality of schooling.

The studies, however, did bring about a better grasp of the problems inherent in assessing the impact of social programs. Even Campbell's broadside, properly regarded as a more differentiated conception of the evaluation process, draws on lessons culled from the earlier investigations.

The Instructional Dimensions Study

The Compensatory Education Study derived from this legacy a clearer awareness of factors impeding the assessment of multiform social programs. The study staff recognized the delineation of appropriate and feasible foci as a crucial step in framing the inquiry. The areas of Title I impact selected for examination were broadened at the same time that the questions addressed within the areas were sharpened more than had been the case in past evaluations.

Four major areas were selected: (1) funds allocation, (2) service delivery, (3) student development, and (4) program administration. The present paper deals with just one of the 35 separate studies that made up the total effort, namely, the Instructional Dimensions Study (IDS). The major research effort in the area of student development, the IDS was designed to evaluate the effectiveness of instruction provided to compensatory education students. Its scope was limited to reading and mathematics instruction at 1st- and 3d-grade levels. More specifically, the study examined the relationship between selected characteristics of instruction and student development. Four dimensions of instructional settings were chosen for close scrutiny: (a) pullout and mainstream instruction, (b) intensity of instruction; (c) content of instruction; and (d) individualized instruction. The summary of findings reported in the Effects of Services on Student Development (NIE, 1977) reveals that overall, the students in the sample registered high rates of achievement. Five sets of classroom processes (opportunity, individualization, instructional events, motivational processes, teacher background) were found to be strongly related to achievement, when taken in combination. By itself, the effect of individualization was not found to bear a significant relationship to student achievement compared with other modes of instruction.

The present paper concerns the classroom processes subsumed under the dimension of individualization. It grew out of the author's participation in a 2-day meeting arranged by NIE, when 40 of the study teachers were convened to discuss and reflect on the findings in light of their knowledge and experience.

For purposes of the IDS, individualized instruction has been defined to conform with the intent of the legislative mandate, as interpreted by the NIE study staff. By singling out the "use of individualized written educational plans for children," Congress was indicating special interest in the performance of programs using this practice. The definition ultimately used was more inclusive, encompassing a set of strategies believed descriptive of individualized instruction as it is typically implemented in schools.

Classrooms were rated on the degree of individualization by the use of a combined assessment of selected classroom processes, primarily:

(1) The assignment of specific learning objectives or activities to individual children

(2) The use of diagnostic and prescriptive activities

(3) The existence of alternative learning paths and sequencing for individual children

(4) The use of individual or small-group pacing

In this interpretation, the use of tests became the sine qua non of individualization. The overall classroom score was heavily weighted toward individualization when tests were used for setting objectives, placing students, assessing progress, and confirming mastery.

The strategy of using features of the instructional process or the classroom setting as predictor variables represented a step forward from evaluations based on global program comparisons. The latter approach almost invariably found within-program differences equal to or greater than between-program differences in outcome, owing in part to program variations across localities, schools, and classrooms.

Considerable and creditable work went into specifying the instructional process characteristics that would signify individualization (Cooley and Leinhardt, 1975). The instructional techniques selected fell under the general rubric of diagnostic/prescriptive instruction, a model now receiving a good deal of attention. Individualized instruction was defined to represent this general model, in the expectation that the claims and counterclaims made in its behalf would be adjudicated by sound evidence.

A Short History of Individualized Instruction

Although differences in learners had been recognized before the advent of mental measurement, the greatly increased use of intelligence and achievement tests after World War I, which formally documented diversity among students, confronted the schools with an open challenge to acknowledge and respond to the heterogeneity of their population. As early as 53 years ago, the National Society for the Study of Education devoted its yearbook to the topic of "Adapting the Schools to Individual Differences." In the introduction to the volume, Washburne (1925) reported:

> It has become palpably absurd to expect to achieve uniform results from uniform assignments made to a class of widely differing individuals. Throughout the educational world there has therefore awakened a desire to find some way of adapting schools to the differing individuals who attend them. This desire has resulted in a variety of experiments. (p. 1)

Individual differences in learning, in fact, had been noted long before the advent of public schooling. An early example of individualized instruction, dating from the first century, can be found in the Jewish Passover service, which commemorates the exodus of the Hebrews from Egypt. In the ritual of retelling the story of oppression, resistance, and rescue through the parted Red Sea into the Sinai Desert, the leader of the service is instructed to suit the telling of the story to the children who are listening. Four kinds of children are identified: the wise child, the irreverent child, the simple child, and the child who does not know how to ask any questions. The prescription of how to vary the telling of the story is thus based on a diagnostic process that analyzes the nature of questions children ask.

More recently, the prevalence of learner differences was again noted by Gage (1967), who raised some discomfiting questions:

> . . . Learners do differ in ways relevant to their ability to profit from different kinds of instruction, content, incentives, and the like. Almost by definition, instruction adapted to these individual differences should be more effective.
>
> If so, why has not the evidence from attempts to individualize instruction yielded more dramatic results? Why are not the mean scores on achievement measures of pupils taught with due respect to their individual needs and abilities substantially higher, in unmistakable ways, than those of students taught in the conventional classroom, where everyone reads the same book, listens to the same lecture, participates in the same classroom discussion, moves at the same pace, and works at the same problems?

Reflecting on the same questions, Glaser suggests that the effects of individualization are masked by the lack of detailed information about the relationships between instructional methods and individual learning capabilities. He identifies a set of necessary conditions for the implementation of individualized instruction (Glaser, 1967):

1. The conventional boundaries of grade levels and arbitrary time units for subject-matter coverage need to be redesigned to permit

each student to work at his actual level of accomplishment in a subject-matter area, and to permit him to move ahead as soon as he masters the prerequisites for the next level of advancement.

2. Well-defined sequences of progressive, behaviorally defined objectives in various subject areas need to be established as guidelines for setting up a student's program of study. The student's achievement is defined by his position along this progression of advancement.

3. A student's progress through a curriculum sequence must be monitored by adequate methods and instruments for assessing his abilities and accomplishments so that a teaching program can be adapted to his requirements.

4. Students must be taught and provided with appropriate instructional materials so that they acquire increasing competence in self-directed learning. To accomplish this, the teacher must provide the student with standards of performance so that he can evaluate his own attainment, and teaching activities must be directed by individual learner accomplishment.

5. Special professional training must be provided to school personnel so that they can accomplish the evaluation, diagnosis, and guidance of student performance that is required to organize instruction for individualized learning--as contrasted to the total-class management of learning.

6. The individualization of instruction requires that the teacher attend to and utilize detailed information about each student in order to design appropriate instructional programs. To assist the teacher in processing this information, it seems likely that schools will take advantage of efficient data processing systems. (p. 3)

The origins of the definition of individualization used by NIE are clearly discernible in this outline. In fact, the index devised to rate the degree of individualization in classrooms incorporated its essential features. In the classrooms rated high on individualization, then, the requirements posited by Glaser can be assumed to have been met, at least to a practicable degree. Yet, individualization was not found to benefit compensatory education students more than other types of instruction. Apparently, not even under these conditions is the expected superiority of individualization readily demonstrable.

In an effort to gain a better understanding of the IDS results, NIE convened 40 teachers to review, discuss, and help interpret the major findings of the study. The teachers' perspectives are highlighted in the next section.

The Teachers' Views

The assembled teachers reflected the spread of practices in the total study sample, which was considerable. The ratings of the teachers' rooms on the index of individualization were not known to the conference participants, nor were the teachers aware of their own rating. The comments and judgments about the proffered model thus could not be related to the teachers' ratings, although they could be linked to specific practices described.

The discussions testified to the difficulty of isolating an instructional dimension for analytic focus. Although individualization was the topic for the sessions on which these comments rest, other dimensions, such as setting, content, opportunity, etc., regularly entered the discussions.

The exchange of views was hindered by a problem of verbal labels. Individualization, a term lacking a commonly agreed-on referent, nonetheless had positive connotations for the teachers. All believed they practiced it, even if through different approaches and to varying degrees. For the teachers, individualization was a label to which a good deal of personal and experiential meaning accrued. When confronted with NIE's definition --asked, in fact, to replace their own definition with NIE's, elements of which did not overlap, or even ran counter to their own practices--considerable confusion ensued. While talking about the benefits, drawbacks, and classroom practices related to individualization, the teachers inevitably lost grip on the newly acquired definition and responded in terms of their own, longer held understanding of the term. This confusion of tongues notwithstanding, the discussions clarified some of the issues raised by study.

A most important contribution was the teachers' evident immersion in the unadulterated

---

[1] The index was derived from several differentially weighted variables yielding a single number that allowed the ordering of classroom on a linear scale. The teachers were aware of the component variable, but not of how they were combined.

reality of classroom life, which contrasted markedly with the schematized representations to which evaluation studies necessarily reduce it. The teachers' talk, brimming with particularistic details rooted in day-by-day experience, was a sharp reminder of the risks of nongrounded theorizing and of the premature imposition of conceptual schemata on events and settings whose salient dimensions are only partially understood. The discrepancy between the teachers' reality and the desiccated variables used in most studies that aim for generalizable outcomes across a wide range of settings and farflung populations was apparent throughout the meeting.

Accounts of quotidian routines and activities anchored much of the discussion. Although time limits kept teachers from describing programs in depth and fine detail, the diversity of their practices was notable. The specifics of the NIE definition were played off their variegated approaches. Bare bones of variables such as packaging, sequencing, etc., were fleshed out by classroom vignettes. Cutting through the surface diversity of the practices described, one could infer some commonalities in the way teachers construed the dimensions of individualization: teachers found the diagnostic/prescriptive model most easily, if not exclusively, applicable to the use of certain types of instructional materials, specifically those with objectives organized into hierarchies. The model was also seen as more suited to the attainment of concrete, discrete, short-term instructional goals than to long-term, integrative, developmental aims stated in holistic terms.

As the structure of classroom organization was not a pivotal aspect of the NIE definition, it was surprising to realize that, in many of the teachers' judgments, the only organizational arrangements that fitted comfortably with the requirements of individualization were one-to-one tutorials, or situations in which the child worked alone with self-paced materials. Apparently, the other components of individualization so constrained the possible arrangements as to rule out a common form of classroom organization--small groups. This perception had important consequences for the way teachers judged the assets and liabilities of individualization.

Major Themes

The teachers' comments did not fold into neat classifications. Recurrent concerns emerged more as leitmotifs that were threaded through the discussions. Two of these will be briefly elaborated. Both relate to the generic issue of the teacher role: one concerns goal setting, and the other speaks to the place of teacher judgment in the pedagogical process.

Teaching Goals. The breadth and complexity of teaching goals subsumed by the definition of individualization proved to be a thematic concern. As mentioned before, the diagnostic/prescriptive model was seen by most teachers to apply to the teaching of academic subjects, if these were divisible into specifiable units that could be learned in the short term, and their mastery could be confirmed by an available test. Some of the teachers seemed to construe their role as consonant with these goals; i.e., they saw their primary task as helping children acquire the component skills that are assumed to cumulate into higher level abilities. Not surprisingly, more teachers who were not in charge of classrooms held this view, namely, the compensatory teachers who worked with children for limited time periods on a specific subject that was judged in need of remediation. The majority of teachers, however, had a different conception of their role. They regarded the aims and purposes of elementary school teaching as more comprehensive, longer range, and, above all, more interrelated. Many found it difficult, if not unproductive, to isolate academic learning from the broader developmental and social goals they held for the children—goals that guided their instructional decisions and practices in large measure.

One teacher spoke for many of her colleagues when she enumerated some qualities of the learning environment she intended to create:

> I am concerned that children have time to express their opinions, their feelings, that they be listened to, have the opportunity to deal with disagreement, with each others' ideas. Time is also needed for the development of critical skills, for comprehension.

The development of a sound and serviceable self-concept also loomed large in most teachers' agenda. "Self-concept" seemed to be the teachers' shorthand for expressing their concern for the student as a person, for placing the more academic objectives in relation to the learner. Supporting positive self-regard was seen as cutting across all of teaching and, for many teachers, as one of their most important tasks.

Perhaps the most troubling feature of the diagnostic/prescriptive model for the teachers stemmed from their perception that it precluded small learning groups. Teachers were forceful in their arguments that children need to work together. They tended to see their classroom as an organic social unit, not as a loose aggregate of individual children. The development of interpersonal codes, social rules, and reciprocal exchanges was not only valued for itself but was regarded as the necessary context for optimal

cognitive development. To help children achieve a modicum of acceptance and respect from their peers, to enable them to secure a place in the class community was a highly valued, if not always fully realized, priority. Encouraging mutual aid, providing opportunities for children to listen to each other, comparing ideas, and resolving or accepting differences were all subsumed under this general goal and linked to the development of self-esteem. Most teachers were mindful of the relationship between the students' self-concept and their capacity to learn. These compensatory education teachers, who by virtue of their role did not see children in their classroom habitat, dealt with similar concerns as they talked of the efforts they made to present participation in the program as a privilege and the pains they took to protect children from the stigma of being singled out for remediation.

The effects of children working on their own were not seen in negative terms only. A few teachers looked on solitary work as a way to protect children from failure in public and from excessive competition with peers. Others saw opportunities for the development of self-direction and autonomy. The purpose and extent of working alone finally determined teachers' view of the practice; they recognized the need for children to do so, but became concerned when the line between working independently and working in isolation became blurred.

The holistic goals that many teachers shared were not easily absorbed by the diagnostic/prescriptive model; they are not readily segmented into stepped learning objectives, nor is their mastery measurable by standardized instruments. Yet such goals have great salience for teachers, with significant influence on their teaching practice. In part to accommodate these aims, most teachers modified the programs they were implementing. It is important to recall here that the teachers all had had a minimum of 1 year's experience with whatever program they were using and thus a chance to put their own stamp on it. The adaptations ranged widely, most falling short of the teacher whose response to how she shaped the program was "Actually, I modified it by mostly throwing it out." Other teachers described limiting prescriptive/diagnostic teaching to portions of the schoolday and supplementing the prescribed materials and activities.

Teachers adapted the programs in response to individual differences--to take account of children's interests, learning styles, or motivational factors, based on their own judgment as to what would work best.

Teacher Judgment. The topic of teacher judgment was perhaps the dominant theme of the meeting. Teacher judgment is required to transform general aspirations for a class into differentiated goals for individual children. Professional judgment is even more deeply implicated in the way teachers seek to realize their pedagogical aims. This, of course, is the heart of teaching--selecting and providing the means to best support the development of each child. The diagnostic/prescriptive programs, when followed to the letter, do not leave much room for the exercise of teacher judgment. The what of teaching is determined by prescribed objectives, children's needs are determined by diagnostic measures, and the how of teaching is supplied by prescribed materials and activities. The tightness of the prescriptions varied across programs, from the specification of exact steps to be followed and of skill drills and techniques for reinforcing the skills, to programs that provided alternative materials and paths that teachers and students could choose to follow. Again, the teachers whose role was limited to compensatory instruction tended to have fewer options in what and how they taught. One teacher remarked that she sometimes felt like an automaton, so restricted was her teaching domain.

The sound exercise of teacher judgment, however, exacts its own imperatives. The opportunity to observe and interact with students, to "study" them, in effect, is indispensable for acquiring the information needed to respond to children differentially. The diagnostic and mastery tests built into the individualized programs provided an abundance of information about student performance on the skill objectives. Teachers used these to sequence the next objective and to assess the pace of students' progress. The tests, however, did not speak to other prominent teacher concerns. The results did not inform teachers about the nature of the children's interests, their preferred modes of work, what motivated them, what aroused their anxiety, or how they felt about themselves. Several teachers shared the conviction that elementary school instruction requires this comprehensive view of the learner. One of the teachers enumerated part of what she attends to when working with children:

> . . . I observe children to find out how do they learn best, how do they function in a group, do they pay attention all the time or some of the time . . . how are they cared for, how are they dressed . . . to find the way to assure them, do they like to be hugged, or be praised . . . .

Forming such judgments was seen as pivotal to the teaching function. Teachers did not expect

assessment instruments to do it for them, but they did need the opportunity to observe students working with others, to see them function in a variety of settings, as well as to interact with them one-on-one. A close adherence to the routines of a diagnostic/prescriptive program closed off some of these opportunities: it curtailed group activity, the prescribed objectives limited the variety of learning goals, and the segmentation of the curriculum content was not conducive to revealing the students' understandings and thought processes in much depth or scope.

Few teachers, judging by their own reports, implemented the individualized programs by the book. When their situation permitted it, they exercised their judgment and acted on their priorities. Programs were adapted to suit the children, as well as to harmonize with the teachers' preferences, values, and skills. Perhaps the most paradoxical aspect of many individualized programs is the failure to acknowledge that teachers too have interests, preferred approaches, and value perspectives. The discussions provided ample evidence that teachers differ in the ways they construe their role and in the organizing priorities that guide their practice. It was this diversity of perceptions that resulted in the scrupulous implementation of programs by some teachers and their partial or extensive modification by others. Teachers, however, are not always free to determine the course of their own instruction, which leads to varying degrees of overlap between what they feel comfortable and competent to do and what they in fact do. A misalignment between overt teaching practices and the teacher's underlying pedagogical belief system can have unwelcome effects (Bussis et al., 1976).

Contextual Factors. In the foregoing sections, differences in assumptions about educational goals and practices held by the teachers and those underlying the IDS were pointed out in an attempt to begin to understand why individualization did not relate to student gains. In the next section, features of the broad school setting, which emerged during the discussions as likely to affect student achievement, will be described briefly. Although these situational factors were not considered in the IDS, they have plausibility as influential variables and serve to point out once again that similarly labeled teaching practices may mask vital differences.

Typically, compensatory education was "delivered" to the students by someone other than the classroom teacher. In the case of pullout programs, a special Title I teacher, often a reading or math specialist, was responsible. In programs that were mainstreamed, it was a special teacher coming in the room or an aide who was identified as the compensatory education teacher. The teachers attributed considerable importance to the relationship between the child's classroom teacher and the teacher responsible for the compensatory instruction. In schools where students were pulled out from their classrooms, this relationship was especially important, as the classroom teacher was not present during these instructional episodes. The nature of coordination ranged from close collaboration between the teachers, involving joint goal setting and frequent information exchange about the students' progress, to virtually no contact between the two teachers, and therefore no coordination of the children's experience in the two settings. In rooms that mainstreamed the compensatory education program, coordination was less of a problem. When compensatory education funds supported a classroom aide, it was virtually ensured. In cases where special compensatory education personnel visited the classroom to work with selected children, the program tended to be more independent of classroom instruction.

Another source of variation was reported in the way schools scheduled the compensatory education program. In some situations students received compensatory instruction while their classmates had music, art, or physical education, watched films, or participated in other, not strictly academic activities. In other schools, every effort was made to see that the children were not deprived of common class experiences in favor of compensatory education.

Yet another variable mentioned was the degree of parent involvement. Examples were cited of extensive training programs offered to parents in the principles of diagnostic/prescriptive instruction, along with materials and guides urging parents to continue the instruction at home. More commonly, parents were occasional visitors to the school, maintaining differing degrees of contact.

The training of the staff administering compensatory education was another factor mentioned. In schools where special, or auxiliary, staff members were responsible, not only their training, but the status they hold in the school community, and even the location where they carry out their work may be influential factors. As was the case with every other variable, notable differences were reported. In some schools Title I staff were highly trained professionals; in some other schools, they were highly trained nonprofessionals. In some instances, preparation seemed modest. The place of instruction, in the case of pullout programs, varied, most likely in concert with the status of the position. Special rooms,

well-equipped "labs," were the scene of instruction in some schools, while the hallway served the purpose in others.

Schools also differed in the proportion of students receiving compensatory education, which almost certainly reflected the socioeconomic background of the student body--a variable often found correlated with scholastic achievement. The teachers also mentioned curriculum and program relation differences. The approach to instructing Title I students was generally consistent in some schools, while in others more than one approach or program was being implemented.

The quality and substance of collegiality available to teachers was reported to be an important determinant of the teachers' perceptions of their work environment, as was the role of administrative support in maintaining a favorable school climate. As might be expected, the schools differed on these dimensions as well.

Most if not all of these variations in school settings and routines can be assumed to be orthogonal to the degree of individualization, yet their relationship to achievement scores cannot be ruled out.

Coda

Only magical thinking would lead anyone to expect that 2 days of freewheeling discussion with 40 teachers would make the findings of an intricate and ambitious study fully comprehensible, yet the teachers' comments provide interpretive guides not present in the other data.

The IDS speaks to two questions that have preoccupied educational research and evaluation; it focuses on one directly, on the other more obliquely. The relationship beween selected instructional practices and student development was the central study question. The selection of an appropriate strategy for uncovering such links was addressed less directly. These seemingly separate issues ultimately converge, for the assumptions underlying one invariably shape the approach to the other.

First, a comment on evaluation strategy. The conceptualization of individualization as an instructional technique--a sum of a set of instructional practices that could be measured across settings--largely excluded contextual and student- or teacher-related factors from consideration as variables affecting learning outcomes. Such an assumption is only justified if instructional techniques do not interact with qualities of teachers or students, but instead unfailingly produce their effects across settings. The teachers' accounts, which suggest that similar practices are perceived and valued differently by different teachers and in different school environments, cast considerable doubt on this premise.

The design of the IDS was also shaped by the expectation of additive linear relationships between the indicators of individualization, such as pacing and sequencing, and student attainment. The presumption of a simple, regular form of "the more the better" has been challenged repeatedly for psychosocial variables, where curvilinear relations are often found.

Finally, there is the insistent theme of the nature of diagnostic/prescriptive instruction itself and the relationship it bears to student learning. Diagnostic/prescriptive programs consist primarily of a sequential ordering of subject matter, with instructional suggestions on how to deploy the accompanying diagnostic instruments and teaching materials. Such programs may be looked upon as resources and tools provided for the teacher to use at her discretion, or, in contrast, as directives on what and how to instruct students. These polar views, and the gradations between them, were represented in the group of teachers convened, with the majority advocating using the principles and material of the individualized programs as teaching aides in a broader context of pedagogical priorities and practices. Teachers generally assumed responsibility for a far more comprehensive domain than the one typically encompassed by the skill objectives of diagnostic/prescriptive programs. More important, the processes of acquiring such skills were seen as related to qualities of learners and their interactions with modes of instruction--factors that teachers wished to take into account as they mapped out paths to learning for the students. Teachers, in essence, aspired to personalize instructional encounters, not only to individualize them.

The IDS findings indicate that students can attain high test score gains in individualized programs implemented by experienced teachers, but that comparable gains are achieved by students in less individualized programs. The teachers' comments also suggest that the issue goes beyond the question of what works, in achievement test terms, centering finally on the value-laden question of what it is that we want to work--what image of the educational process, what roles for teachers and learners are envisioned and desired. One image of the learning child is of an active, purposeful being who interacts, rather than simply responds to the opportunities and constraints present in any educational setting. The learning process, in short, is

seen as embedded in the person. The parallel to this conception of the learner is found in the perception of the teacher as an active, purposeful decisionmaker who uses the instructional resources available to best realize differentiated goals formulated for the students. There are other visions, of course, that grant less autonomy to both the teacher and the learner. The model of individualized learning used in the IDS is more consonant with the latter perspective.

It is instructive to return to Glaser once more, a full decade after his conceptualization of individualized instruction. In the intervening years the model has evolved and transformed to become adaptive education. Glaser (1977) briefly describes the essential elements of an educational system that is adaptive to the individual:

> It provides a variety of alternatives for learning and many goals from which to choose. It attempts to utilize and develop the capabilities that an individual brings to these alternatives and to adjust to the learner's particular talents, strengths, and weaknesses. Also, an adaptive educational environment attempts to strengthen an individual's ability to meet the demands of available educational opportunities and develop the skills necessary for success in the complex world.

The implications of this vision are far-reaching. It grants the learner active participation in instructional decisions while implying the sustained presence of a teacher who observes, evaluates, and provisions. Such an environment will not result from the implementation of preordained learning objectives and instructional activities designed in the absence of particular learners. The image of adaptive education calls for an investment in professional development rather than prepackaged curriculum plans and materials. It also suggests the legitimization of evaluative judgments about a broad array of learning priorities and outcomes by more informal, teacher-centered procedures than those embedded in standardized assessment techniques.

It is to be hoped that the findings of the IDS will result in the pursuit of these bolder aspirations for schooling.

BIBLIOGRAPHY

Berryman, S., and O.T. Glennan. An Improved Strategy for Evaluating Federal Programs in Education. Draft Working Note 10092-ASE. Washington, D.C.: The Rand Corporation, May 1978.

Bussis, A., E. A. Chittenden, and O. M. Amarel. Beyond Surface Curriculum: An Interview Study of Teachers' Understandings. Boulder, Colorado: Westview Press, 1976.

Campbell, D. Assessing the Impact of Planned Social Change. Occasional Paper #8. Kalamazoo, Michigan: Western Michigan University, December 1976.

Cooley, W. W., and G. Leinhardt. Design for the Individualized Instruction in the Teaching of Reading and Mathematics in Compensatory Education Programs. Final Report. Pittsburgh, Pennsylvania: University of Pittsburgh, Learning Research and Development Center, 1975.

Dyer, H. "Thoughts on the Mournful Numbers." Paper presented at the Annual Meeting of the American Educational Research Association, Toronto, March 1978.

Gage, N. L., and W. R. Unruh. "Theoretical Formulations for Research on Teaching." Review of Educational Research 37(3):358-370, 1967.

Glaser, R. "Adapting the Elementary School Curriculum to Individual Performance." Proceedings of the 1967 Invitational Conference on Testing Problems. Princeton, New Jersey: Educational Testing Service, 1968.

Glaser, R. Adaptive Education: Individual Diversity and Learning. New York: Holt, Rinehart and Winston, 1977.

House, E. The Objectivity, Fairness, and Justice of Federal Evaluation Policy as Reflected in the Follow Through Evaluation. Urbana, Illinois: University of Illinois, Center for Instructional Research and Curriculum Evaluation, February 1978.

House, E., V. Glass, L. McLean, and D. Walker. "No Simple Answer: Critique of the Follow Through Evaluation." Harvard Educational Review 48(2):128-160, 1978.

McLaughlin, M. Evaluation and Reform: The Elementary and Secondary Education Act of 1965/Title I. Cambridge, Massachusetts: Ballinger, 1975.

National Institute of Education. The Effects of Services on Student Development. Washington, D.C.: National Institute of Education, U.S. Department of Health, Education, and Welfare, 1976.

Spady, W. G. "The Impact of School Resources on Students." In W. Sewell, R. Hauser, and D. Featherman (eds.), Schooling and Achievement in American Society. New York: Academic Press, 1976.

Stanford Research Institute. Implementation of Planned Variation in Head Start: Preliminary Evaluations of Planned Variation in Head Start. Stanford, California: Stanford Research Institute, 1971.

Stebbins, L. B., R. G. St. Pierre, E. C. Proper, R. B. Anderson, and T. R. Cerva. Education as Experimentation: A Planned Variation Model, Vol. IV-A: An Evaluation of Follow Through. Cambridge, Massachusetts: Abt Associates, Inc., 1977.

Wargo, M., G. Tallmadge, D. Michaels, D. Lipe, and S. Morris. ESEA Title I: A Reanalysis and Synthesis of Evaluation Data from Fiscal Year 1965 Through 1970. Palo Alto, California: American Institutes for Research, 1972.

White, S. H., M. C. Day, P. K. Freeman, S. A. Hartman, and K. P. Messenger. Federal Programs for Young Children, Vol. II: Review of Evaluation Data for Federally Sponsored Projects for Children. Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1973.

WHAT DO WE KNOW ABOUT INDIVIDUALIZED INSTRUCTION?

Jane A. Stallings
SRI International
Menlo Park, California

Introduction

The intent of this paper is to describe the origin of compensatory education programs and to show how individualized instruction (or education for individuals) evolved from compensatory education programs. The paper was prompted by the author's participation in a 2-day conference arranged by the National Institute of Education (NIE) to consider the findings from the Instructional Dimensions Study (IDS). This is one of 35 studies undertaken by NIE to evaluate the effectiveness of compensatory education. Also attending the conference were 40 teachers who had participated in the IDS. My particular charge was to report the teachers' responses to the findings regarding individualized instruction. The study showed that students in programs rated high on individualization did not have more progress than ones in programs rated lower. I have attempted to contrast the IDS definition of individualized instruction with the definitions offered by teachers and the definitions used in other research. Findings from other research on individualized instruction are compared with the IDS findings.

Compensatory Education

In response to public demands, Congress has allocated vast sums of money to improve the life chances of minority groups. It was not difficult to see that failure began for minority people in the first grades of school. National standardized testing programs showed that the disadvantaged populations scored well below national norms. In an effort to reduce educational inequities, Congress distributed funds through the states in the form of Title I - Title VII programs. Much of this funding has been directed toward the education of the economically disadvantaged, and has taken account of the best professional advice educators could offer.

Because minority children had so often scored below the norm on standardized tests, educators and Congressmen alike agreed on one point: traditional curricula, materials, and methods did not meet the needs of minority populations. No one really knew under what conditions the economically and educationally disadvantaged would prosper, so that the education programs finally funded represented a wide range of educational approaches. Each approach was some educational theorist's answer to meeting the needs of the educationally and economically disadvantaged. This method of funding was not free of problems, since some experiments were not as effective as others, but it has had the great advantage that, over more than a dozen years of Congressional support, these experimental education programs have produced a large body of information and studies on the effects of compensatory education.

Individual Abilities and Education

Work by Guilford (1967) on individual abilities released educators from thinking in the single dimension of intelligence quotients and spurred interest in the conditions that would promote learning for students with differing abilities. In the 1960's and 1970's a considerable number of studies were conducted to see whether students with particular aptitudes or abilities prospered in particular learning environments. For example, how did students varying from low to high visual sequencing ability prosper in learning environments using phonic or whole-word reading methods? This type of study was called Aptitude Treatment Interactions (ATI). Cronbach and Snow (1977) have summarized the findings from a sizable body of the research.

It was this thinking about individual abilities and treatments that prompted educational theorists and practitioners to vary from the traditional methods of lecture, drill, review, and test. In the 1960's, English Infant Schools in Leicestershire County, U.K., were experimenting with open classrooms where children selected their learning tasks for the day from a variety of learning centers. Children were believed to be innately curious and, given a wide selection, it was believed that every child would choose what he or she needed for a sound educational diet. Many American schools subsequently followed some variation of this pattern, and many schools in the

late 1960's and 1970's followed the open-classroom approach. How the curriculum was managed within the open classroom varied considerably.

With the vast sums available from the Government for research in education, many ideas about what would work best with low-income, educationally disadvantaged children could be tried. Because reading was considered the springboard to other knowledge, the teaching of reading became the focal point of much experimentation. New methodologies based upon guesses and experience of researchers and practitioners were set into practice. These included color-coded reading, behavior modification, experience charts, task analysis, phonetic analysis, the Slingerland method, the Fernald method, the Sullivan method, student contracts, and many combinations of the above. Book companies' explorations seemed limitless. Many new phonetic analysis series and programed materials hit the market. The old "Dick and Jane" books were replaced by books showing children with darker skins and showing more pictures of inner-city living and less suburban and rural living. There was enough money to try everything, and it seems as though everything was tried.

Evaluation

Having funded educational reform, Congress then sought to learn whether or not the additional dollars were improving the education and thus the life chances of economically disadvantaged children. To answer this question, several large evaluations were mounted. For the most part, the findings from the first studies conducted in the late 1960's and early 1970's were disappointing. The Westinghouse Ohio report of Head Start found that gains students made in preschool were no longer evident at the end of 1st grade. Coleman and the many others who reanalyzed his data (Mosteller and Moynihan, 1972; Jencks, 1972) found no relationships between progress students made and what the schools provided in terms of services and programs. In a more extensive review, Spady (1973) also reports mixed results from studies focusing on the relationship between measurable school resources and student learning. However, all of these studies measured differences between schools, not the variation of treatment within classrooms within schools. (Recent studies by Stallings et al. [1977, 1978] have found as much variance within schools as among schools, whether on reading gain scores or on pretest and posttest scores.)

Researchers and practitioners close to the educational innovations being implemented were convinced that these large-scale evaluations were overlooking the most important variables--that is, what was occurring in individual classrooms. In the early 1970's, Brophy and Evertson, Brophy and Good, Soar, and Stallings sharpened their focus and began to study specific elements of education programs as they were carried out in classrooms. Such classroom research required new methods and new statistical techniques. Variation in classroom environments and instructional processes had to be documented in quantitative forms that could be used in statistical analysis.

Several systematic observation systems were developed. Considerable effort was expended toward reliably recording the materials being used, the grouping arrangements in classrooms, the instructional patterns of teachers, and the behavior of children. With this documentation in hand, it was then of interest to see how children prospered in the various environments. Previous studies of schools had considered student scores on standardized achievement tests only. While success in reading and math remains of primary interest, measures of social, emotional, and other cognitive development were used in some studies to provide a more holistic estimate of student growth.

Several such observational studies relating the components of educational programs to student outcomes were conducted (Soar, 1973; Stallings, 1973; and Stallings and Kaskowitz, 1974) in the evaluation of the U.S. Office of Education's National Follow Through Planned Variation program. The Follow Through program funded the implementation of 22 models of education, based upon a wide spectrum of psychological and educational theories and practices. The questions of interest to the government were: (1) Were the programs as specified actually implemented in the classrooms? (2) If so, how did the students prosper in the programs?

A study (Stallings and Kaskowitz, 1974) of seven Follow Through models in 1st and 3d grades indicates that the model programs were being implemented in many of the classrooms. However, of even more importance was the finding that particular components of programs were significantly related to specific types of student growth. Many children seemed to prosper in reading and math in classrooms where teachers were very directive, used sequenced, self-pacing programed materials, provided extensive drill and practice during reading and math periods, provided immediate feedback for right or wrong answers, and allocated approximately 50% of the schoolday to these academic activities. These particular components were used in models based upon the theory of behavior modification. Three models based upon this theory--the University of Pittsburgh's Individualized Early Learning Pro-

gram, the University of Kansas's Behavior Analysis Approach, and the University of Oregon's Engelmann/Becker Model for Direct Instruction--used a variety of organization plans. In one case teachers worked with one child at a time, in the other two they worked with small groups. Two of these models provided verbal feedback; the other model included tokens with the feedback. One model had group choral response; the other two had individual response. The study identified the more global organization and management variables. Each of the three models specified above also carried out very careful evaluations of their student week-to-week and year-to-year progress in reading and math.

Other findings (Stallings and Kaskowitz, 1974) indicated that students in more open classroom models, where students were encouraged to experiment with materials and learn through inquiry, obtained higher scores on a test of nonverbal problem solving. These children also asked more questions about subject matter and more often worked together on joint projects. The school-level evaluation of Follow Through models carried out by Anderson et al. (1977) found no consistent model effect, but this evaluation did not consider classroom process data, only school-level test scores for each grade.

Instructional Dimensions Study

A summary of Title I evaluations by McLaughlin (1975) suggested that most of these evaluations were unproductive and could not guide Federal policy. Therefore, the National Institute of Education was mandated by Congress in 1975 to conduct a study of purposes and effectiveness of compensatory education programs. These programs were supported by Title I of the Elementary and Secondary Education Act (ESEA) and state funding.

To avoid the pitfalls of the large-scale studies, the NIE staff selected four major areas for study: funds allocation, service delivery, student development, and program administration. These areas were then sharply focused and a total of 35 separate studies were defined.

The present paper deals with one dimension of the student development area, the Instructional Dimensions Study (IDS), which was designed to evaluate the effectiveness of instruction provided to students in compensatory education programs. Based upon prior research, expert advice, and judgment, four dimensions of instructional settings were selected for study: pullout and mainstream instruction, individualized instruction, content, and intensity of instruction. The dimension this author was requested to consider is individualized instruction.

Congress had indicated an interest in the "use of individual written educational plans for children." The NIE Study staff interpreted this as an interest of Congress in the performance of students in classrooms using this practice. The long step that had to be taken by the IDS staff was to define what would count as an individualized program. Classrooms were to be rated on the degree to which they practiced individualization by the following criteria:

- The assignment of specific learning objectives or activities to individual children
- The use of diagnostic and prescriptive activities
- The existence of alternative learning paths and sequencing for individual children
- The use of individual or small-group pacing

This classroom-level analysis of process was a great step forward. Most previous studies of Title I had addressed student achievement only and usually at the school level.

Once the IDS started, the definition of individualized instruction was narrowed so that teachers were rated on instructional techniques found primarily in the behavior modification models previously described in this paper. These diagnostic/prescriptive, direct-instruction techniques have received growing acceptance recently so that it is easy to understand why the IDS staff altered the previously stated definition of individualized instruction to the following:

- Clearly stated behavioral objectives which take each student in small steps through curriculum materials
- Individual diagnosis and prescription
- Individual or small-group pacing
- Structured sequential instruction

These were further reduced to four main ideas: establishing behavioral objectives, diagnosing/prescribing, individual pacing, and individual sequencing.

Thus, in the opinion of this author, the definition of individualized instruction became quite limited. Diagnosing and prescribing did not begin to encompass the abilities identified by Guilford in studying individual abilities. Teachers

were given a rating for individualized instruction based on the extent to which they provided structured sequential materials through which students could go at their own rate to meet the objectives which were finally to learn to encode, decode, and comprehend written material.

Teacher Reports

The teachers at the conference were also troubled by the IDS operational definition of individualized instruction. There were several interpretations of the terms. Most, but not all, teachers believed that individualization required working with one child at a time while allowing all children to progress through the same reading program at their own rates. Other teachers felt they provided individualized instruction within small groups. Diagnosing for some teachers meant diagnosing learning style preferences (auditory, visual, or kinetic), while for other teachers it meant diagnosing encoding and decoding skills. Some teachers diagnosed personality styles; i.e., some children prefer to work alone, others can't work alone and need peer support; some need considerable praise and support, others are intrinsically motivated; some need touching and holding, others do not. Thus, when teachers say, "Yes, of course I diagnose and prescribe," there is a wide difference in what actually occurs.

Teachers also varied on how much of the day they spent in individualized instruction. Some said they used it only for reading, for which they allocated approximately an hour a day. Other teachers said they worked most of the day on a one-to-one basis with their students. Others said the amount of time spent on a one-to-one basis should be more for beginning 1st graders and less for 3d graders, who do not need as much individual help.

For some, the individualization took place only in a pullout program, where the child is pulled out of the classroom for instruction. In this system, the ratio of students to adults was usually 10:1 or lower. Pullout programs varied in their structures from formal, well-equipped laboratories, some with computer terminals, to one aide working with one child in the hallway or in a closet. Also, the amount of time children spent in pullout programs varied. Some children spent an hour every day in the laboratory, for a period of 6 weeks; then a new group would go to the laboratory. In other cases, children would go to the laboratory for as little as 20 minutes a day twice a week.

These kinds of variances in the implementation of an individualized program, if not accounted for, could introduce substantial error into the analysis and might explain in part why no more effect could be attributed to individualized instruction.

The teachers were frank in offering their opinions about the usefulness of individualized instruction. Most teachers felt it was too much work, but if there is enough space, if good aides are available, and if the ratio is then 10 children to 1 adult (and if the curriculum is self-teaching), then it is possible to individualize for all children. But even with those conditions, many teachers felt that it was easy to get consumed by the programed materials. One teacher said, "When you have to hurry from one child to the next to be certain that all children get their instruction and feedback, it is difficult to get to know the children and see them holistically."

Some teachers felt that children learn important things from each other--that they need to learn to listen to each other and this doesn't happen when they all work alone. When discussing the idea of a story with a group and developing the vocabulary in the process, all children hear the discussion and have more opportunities to learn. Skilled teachers can involve all of the children in the discussion by distributing questions appropriate to each child's level. This is another way of individualizing, but one that leads to a spirit of cooperation. Other teachers reported that children working alone in workbooks at their own pace tend to worry about who is ahead and who is behind, and the faster children tend to brag about what page they are working on. Although teachers may attempt to focus each child on his or her own progress and not on the page number, this is hard to do.

Opinions of teachers also varied on the usefulness of pullout programs. Some teachers felt that sending children from the classroom for instruction was disruptive and posed problems when trying to fit the children back into the classroom. Scheduling these departures was difficult; some children would then miss other activities, such as music, art, or physical education. Some teachers reported that the work children did in pullout programs was not coordinated with what they did in the classroom.

Other teachers felt the pullout programs were a godsend. The children had fun in the laboratories, where they could use new materials, play games, or use computers, and have the total attention of an adult who knew how to assist them on their level. Meanwhile, the classroom teacher could attend to the needs of fewer children.

The advantage most often voiced for individualization was that children do not have to go too fast or too slow; they work at their own speed and at their own level. Some teachers felt that the very slow children do not get embarrassed publicly when the teacher works with one child at a time, since all of the feedback offered is private. The frequent diagnostic tests of some curricula were

also seen as an advantage. This process allows each child to move on to the next sequence of material when it is clear that the present material is mastered. The programed curriculum materials help to take the guesswork out of their teaching.

Such materials may help provide a consistent program much needed by a highly mobile society. Children moving from one school to another can take their programed materials with them. They would not have to wait until the new teacher figured out where they should start in a new reading program.

In summary, the teachers' points of view regarding how instruction should be individualized were varied. No one position could claim dominance.

IDS and Other Research on Individualized Instruction

Although the IDS found that overall the children in the study gained in reading, the children in classrooms rated high in providing individualized instruction (objectives, pacing, sequencing, diagnosing/prescribing) did not gain more than did children in classrooms rated low on this variable. Individualized instruction is often organized as individual instruction--one adult works with one child at a time. Individual instruction does not mean that an appropriate methodology has been selected for each child; still, it is the variable thought to be most essential by many teachers and program planners. Thus, several studies of elementary and secondary classrooms have looked at this classroom management style. These findings indicate that in classrooms where the primary strategy is to have one adult work with one student at a time, the gain for the total class is less then in classrooms where adults work with groups (small or large). In a study of 48 California 3d grades in the Early Childhood Education Program, a subset of 12 classrooms were matched for entry-level reading scores. Children in six of the classrooms made significantly more gains in reading than did children in the other six classrooms. From an analysis of variance of the two groups, we learned that teachers and aides in the gain classrooms were working primarily with small groups or the total group; teachers and aides in the no-gain classrooms were, most often, working with only one child at a time. (Working with one child at a time occurred 2.5 times as often in the no-gain classrooms as it did in the gain classrooms.) Students worked alone more than three times as often in the no-gain classrooms as they did in the gain classrooms. Also, there was three times as much misbehavior in the no-gain classrooms.

This suggests that when students are left on their own to do seatwork for longer periods of time, or as they wait for their turn with the teacher, they may not get instruction and support as often as they need it and may digress into negative, or at least off-task, behavior. These findings were similar in the SRI Follow Through observation evaluation by Stallings and Kaskowitz (1974) and in the SRI study of teaching basic reading skills in secondary schools by Stallings et al. (1978).

Several classroom process studies indicate that students spend from 40% to 70% of their time working alone (McDonald, 1975; Good, 1977; Stallings and Kaskowitz, 1974). In classrooms where teachers most often deal with one student at a time, the time alone increases. It becomes of utmost importance then, to find how to increase the time students are actively engaged in learning when they are left alone to do seatwork.

How classrooms are managed is a critical problem that has not received sufficient attention. Indicators of poor management include a high rate of student misbehavior. Studies of both elementary schools and high schools report that misbehavior has a strong negative correlation with achievement. Thus, it is important for teachers to provide adequate supervision and guidance, so that children will stay engaged with their work, and it seems they can manage this best when working with the total group or small groups.

Rosenshine (1977) summarizes the findings on grouping in the following way:

> The studies of primary grade classrooms point to the need for adult monitoring and supervising of student activities. Stallings and Kaskowitz Follow Through Study (1974) found that time spent working with only one or two students, was negatively related to class achievement gain, whereas, time spent working with small groups (three to seven students) or with large groups was consistently positively rated to achievement. Likewise, Soar (1973) discovered that when students worked in groups under adult supervision, correlations with achievement were positive and often significant. On the other hand, when small groups met without an adult, correlations between this grouping pattern and achievement were negative and often significant. A simple fact may be inferred from the studies cited: given that many students do not engage in on-task behavior unless a teacher or another adult is monitoring their academic activities, the use of large group settings allows for more adult supervision. Although many educators advocate working with one or two, such an arrangement precludes adequate supervision for the remaining children. As a result, most of the children have less academic engaged time.

It would be interesting to know whether in the IDS there was a higher rate of off-task behavior and less gain in classrooms where there was more time spent with one child, as found in these other studies.

Possible Explanations for Why the IDS Found No Difference in Effect upon Reading Between Classes Rated High or Low on Individualized Instruction

The final IDS definition of individualization used for rating teachers does not specify who the teacher works with--one child, small group, or total group. It includes behavioral objectives, individual pacing, individual sequencing, diagnosing, and prescribing. Nevertheless, most teachers interpret individualization as working with one child at a time, and this management system has been found (as noted in other studies) were not. They were working with small groups or large groups, and to my way of thinking, that is not individualization." Therefore, they were not rated high on individualized instruction by this supervisor, even though they might have used sequential materials with groups of students who established their own pace. Another explanation of the lack of difference in progress between the two groups was suggested by a teacher who thought that many teachers placed in the low individualizing comparison group might have been individualizing effectively in their own way but did not meet the IDS criteria.

What Have We Learned and Where Do We Go from Here?

This study has contributed to our understanding that the definition the IDS offered of individualized instruction (diagnosing encoding and decoding skills and placing students in sequenced programed materials, where they are allowed to progress through the materials at their own rate and finally meet the objectives of being able to encode, decode, and understand written materials) is not sufficient to explain why some students prosper in an environment rated high on these variables while others do not.

Although it is understandable, it is unfortunate (in this author's opinion) that the definition of individualized instruction was limited to the instructional techniques used in the structured diagnostic/prescriptive models based upon behavior modification theory. These received considerable good press for improving the reading scores of some children, but to call the practices of diagnosing encoding and decoding skills and of pacing students through the same sequence of material "individualized instruction" is shortsighted. This does not seem to be what was meant by the Congressional interest in programs that "use individualized written educational plans for children."

Children in any classroom have many aptitudes and limitations; no one method is likely to work for all students. The National Reading Council reports that the 20% of students with reading problems are not aided by present classroom methods and that the percentage of students with reading problems who are failing remains the same today as 20 years ago. Clearly there is a need for a more extensive diagnostic program so to be ineffective in promoting academic achievement. Such a management system could wash out the good effects of individual pacing, sequencing, and so on.

The findings reported from the research do not suggest that individualization, as defined in the IDS, is ineffective; the findings do have implications for classroom organization and management of time.

Other portions of the IDS definition of individualized instruction are found in studies on research in teaching by Brophy and Evertson (1974), Good (1978), Stallings and Kaskowitz (1974), and Stallings et al. (1977, 1978). These studies find positive correlations between reading achievement and specific components of directive teaching. These components include direct questioning, answering, and feedback procedures. The children are tested frequently and placed accordingly in programed materials. The materials are carefully sequenced, and all children either show mastery of the materials on criteria tests or are cycled back through the sequence. For the most part, children are grouped by ability level, and instruction, drill, practice, and feedback are provided to the group. Children may move from one group to another if their achievement surpasses that of the group. Findings from these studies suggest that in classrooms where significant reading gain is made, approximately 50% of the schoolday is allocated to academic study of the type described. Thus, if the amount of time spent in reading is not accounted for, the value of the direct instruction or this individualized approach may be lost. All of these factors are undoubtedly linked together, and when they are dissected to consider the impact of each separate part, the effect of the whole is lost.

In the IDS, the definition of individualization had similar component names as those in direct instruction, but the actual events occurring in classrooms with those labels may be quite different from the events identified in the research under the same label. That is, the lack of effect the IDS found in the classrooms reported

to be more individualized can be understood in light of the different definitions teachers offered for sequencing, pacing, diagnosing, and prescribing. Teachers' self-reports and supervisor's and observers' reports were not consistent. One supervising teacher said, "The teachers I supervise said they were individualizing, but I said they

that perceptual, motivational, emotional, and neurological problems can be identified. Many schools provide only reading encoding, decoding, and comprehension diagnostic data.

Perhaps treatment for reading problems should include eye exercises and kinesthetic activities as well as drill and practice on short sequences of information. Given particular abilities, we still know very little about what is optimal for some children to learn, and particularly how they learn to read. Thus, I think we need to move down one more level: we have moved from the school level to the classroom level over the past decade, and now it is time for studies where individual students are the level of analysis. Building upon work described by Cronbach and Snow, well-designed aptitude treatment interaction studies are needed.

It would be unwise to conclude from the IDS that individualized instruction is not good for students. Children who receive a few minutes of individual instruction while the rest of the class raises havoc may not have the same opportunity to progress as children taught in small groups, but individualized instruction aimed at meeting the needs of individual children whether in a group setting or in a one-to-one setting has not been tested in a way that would show whether it works or not.

Approximately 80% of the students probably did make progress under the approaches used. The 20% who did not may have had other problems that went undiagnosed, and thus untreated, in the IDS. We do not yet know who thrives where--Benjamin Bloom predicts that 99% of the students in school can learn if given appropriate instructional opportunities. It is the job of the teacher and researcher to provide the appropriate instructional opportunities for each child. This study is a step in the right direction, but there is still a long way to go until each child can progress in school and obtain his or her right to a functional education.

BIBLIOGRAPHY

Anderson, R. B., et al. Education as Experimentation: A Planned Variation Model, Vol. IV-D. Report No. 76-196D to the U.S. Office of Education under Contract No. 300-75-0134, April 1977.

Bloom, B. S. Human Characteristics and School Learning. New York: McGraw-Hill Book Company, 1976.

Brophy, J., and C. Evertson. Process-Product Correlations in the Texas Teacher Effectiveness Study: Final Report. Austin, Texas: Research and Development Center for Teacher Education.

Cronbach, L. J., and R. E. Snow. Aptitudes and Instructional Methods. New York: Irvington Publishers, Inc., 1977.

Good, T. L. The Missouri Mathematics Effectiveness Project. Columbia, Missouri: School of Education, University of Missouri at Columbia, March 1978.

Good, T., and J. Brophy. "Analyzing Classroom Interaction: A More Powerful Alternative." Educational Technology 11:36-41, 1971.

Guilford, J. P. The Nature of Human Intelligence. New York: McGraw-Hill Book Company, 1967.

Jencks, C. Inequality, a Reassessment of the Effect of Family and Schooling in America. New York: Basic Books, Inc., 1972.

McDonald, F. J. "Report on Phase II of the Beginning Teacher Evaluation Study." Journal of Teacher Education, Vol. 27, No. 1, spring 1976.

McLaughlin, M. W. Evaluation and Reform: The Elementary and Secondary Education Act of 1975, Title I. Cambridge, Massachusetts: Ballinger, 1975.

Mosteller, F., and D. Moynihan (eds.). On Equality of Educational Opportunity. New York: Vintage Books, 1972.

Rosenshine, B. Primary Grades Instruction and Student Achievement Gain. University of Illinois, 1977.

Soar, R. Follow Through Process Evaluation, 1970-1971. Gainesville, Florida: Institute for Human Development, University of Florida, June 1973.

Spady, W. G. "The Impact of School Resources on Students." In F. N. Kerlinger (ed.), Review of Research in Education, Vol. 1. Itasca, Illinois: F. E. Peacock Publishers, Inc., 1973.

Stallings, J. Follow Through Program Classroom Observation Evaluation 1971-1972. Menlo Park, California: Stanford Research Institute, 1973.

Stallings, J., and D. Kaskowitz. Follow Through Classroom Observation Evaluation 1972-1973. Menlo Park, California: Stanford Research Institute, 1974.

Stallings, J. et al. Early Childhood Education Classroom Evaluation. Menlo Park, California: SRI International, November 1977.

Stallings, J. et al. A Study of Basic Reading Skills Taught in Secondary Schools. Menlo Park, California: SRI International, January 1978.